Matthew Bain, MSc

Data Scientist | Tableau, Databricks | Neuroscience Researcher Turned AI 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.
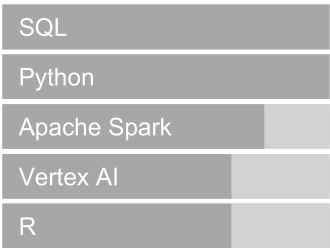
[LeetCode] [Kaggle] [Tableau] [Streamlit] [Hugging Face]

💼 Professional Experience

Present | Jan 2024

Open Source Developer
Independent 📍 Toronto, ON

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

Oct 2023 | Jun 2023

Data Scientist
VocaLinks Inc. 📍 Toronto, ON

- Created and communicated market breakdowns and technical tutorials (R Markdown, Python) on 4 AI-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

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 infrastructure
- 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)

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

Apr 2021 | Sep 2018

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

Contact

- 🗺 Ontario
- 🔗 Personal Website
- GitHub
- in LinkedIn

Core tools

- SQL
- Python
- Apache Spark
- Vertex AI
- R

Platforms

- TensorFlow
- Tableau
- GCP
- Databricks
- GitHub Actions

Key competencies
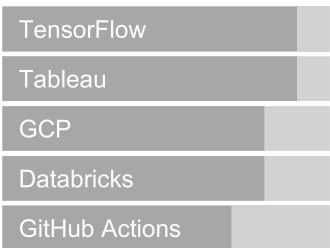
- Model Tuning
- ML Pipelines
- GenAI
- CI/CD
- Bayesian Statistics
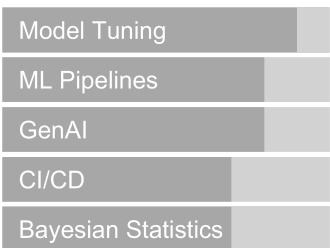
Last updated on 2025-04-11.

Apr 2018
|
Apr 2015

Head of Finance

Lorne Park Landscaping

📍 Mississauga, ON

• 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

Dec 2023
|
Sep 2021

BSc (Honours), Mathematics and Statistics

McMaster University

📍 Hamilton, ON

Aug 2020
|
Sep 2018

MSc, Neuroscience [Thesis]

University of Western Ontario

📍 London, ON

Apr 2018
|
Sep 2014

BSc (Honours), Neuroscience

McMaster University

📍 Hamilton, ON

🏅 Certifications

Mar 2027
|
Mar 2025

Databricks Certified Data Engineer Associate [Certificate]

Databricks

Feb 2027
|
Feb 2025

Tableau Certified Data Analyst [Certificate]

Tableau

Expected May 2025

Professional Machine Learning Engineer

Google Cloud Platform (GCP)

Expected May 2025

Machine Learning Specialization [Certificate]

Stanford University

Expected May 2025

SQL Basics for Data Science Specialization

UC Davis

🔀 Selected Projects

Present
|
Jan 2024

datopy (Data Management Python Package) [Documentation]

• Implemented and maintain a package for handling unstructured data, providing a simple interface for data modeling, extraction, validation, and building ETL pipelines (Pandas, PyTest, Pydantic, GitHub Actions)

Present
|
Apr 2024

mlvizz (Machine Learning Data Visualization Package)

• Build Python interfaces for efficient, intuitive model selection, tuning, inspection, and ML pipelines through modular, object-oriented designs with built-in data validation (Scikit-learn, TensorFlow, ArviZ, PyMC, SciPy)

Present
|
Apr 2024

statvizz (Statistical Data Visualization Package)

• Build Python extensions unifying Pandas data summarization with Seaborn statistical plotting functionality for a unified, intuitive, fully transparent statistical plotting interface (Pandas, Seaborn, SciPy, Matplotlib)

mathvizz (Mathematical Visualization Package)

Present
|
Jul 2024

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

One-stop Superstore Dashboard [Tableau]

Feb 2025
|
Feb 2025

• 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)

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

Dec 2024
|
Nov 2024

• 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)

Nurse Staffing Analysis and Recommendations [Website]

Aug 2024
|
Jul 2024

• Analyzed nursing data using Bayesian models and ARIMA forecasting, revealing regional asymmetries and 90% predictive accuracy up to one month in the future. (statsmodels, Pandas, Scikit-learn)

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

Jul 2024
|
May
2024

• 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)