

David Burton

New York, NY | david@databurton.com | databurton.com | linkedin.com/in/databurton | github.com/Burton-David

Independent data scientist and quantitative developer based in New York. Two-year-old algorithmic trading system on SPY options. Runs live capital. Pays the bills. Collaborating on Alzheimer's research with the NACC dataset (under peer review). Developing computational biology tools for an RNA interference biotech startup. Former international math teacher. Purdue and Indiana University.

EXPERIENCE

Quantitative Developer — Atlas Trading System (Personal)

New York, NY • 2024 – Present

- Built and operate a production algorithmic trading system on SPY options managing personal capital — primary income source
- ML ensemble models with adaptive weighting, 2.48M training data points, sub-millisecond feature calculation
- Event-driven architecture processing thousands of market events per second with 99.5% uptime
- Custom market microstructure indicators, walk-forward cross-validation, automated circuit breakers and risk management

Data Science Consultant — Independent Practice

New York, NY • June 2024 – Present

- Collaborating with University of Kansas neurology department on peer-reviewed clinical research using NACC dataset (31K+ participants), applying survival analysis and statistical modeling
- Elm Dirt LLC: Developed computational biology platform for agricultural biotechnology, implementing sequence analysis algorithms and bioinformatics pipelines for USDA grant submission
- K9Cakery: Built ML-driven inventory optimization system reducing stockouts by 20% and improving demand forecasting accuracy
- Developed ResearchAssistantAgent using LangChain SDK with custom tool calling, FAISS vector search, and multi-source API integration

Data Scientist — BTL Industries

Remote • February 2023 – June 2024

- Built automated lead generation system processing 100K+ business profiles daily using local LLM (Qwen2-7B) and adaptive web scraping
- Engineered data validation pipeline using spaCy for cross-source verification, achieving 95% data accuracy across SQLite operations
- Created B2B data infrastructure enabling sales team to win President's Club with 2x revenue growth (\$30M) vs. teams without the system
- Conducted 3-day on-site intensive workshops for enterprise clients with regular travel for hands-on implementation support

Data Analyst — Prenda

Remote • February 2021 – March 2022

- Built student tracking dashboard integrating 7+ educational platforms into unified PostgreSQL database with pass/fail stoplight system for 10,000+ students
- Designed special education management solution streamlining IEP workflows and improving case handling efficiency
- Developed predictive models using student performance data to flag intervention needs and optimize educational outcomes

Data Analyst — Freelance

New York, NY • September 2019 – February 2021

- Automated grant writing applications using NLP processing, increasing efficiency by 40% and boosting funding acquisition by 200%
- Built predictive models for nonprofit donor behavior analysis and funding opportunity matching across 1,000+ sources
- Created and deployed Instagram chatbot growing active users from 90 to 1,000+ within three months

EDUCATION

Google Cloud Professional Data Engineer Certification • 2024

Flatiron School — Data Science Program • 2019

Indiana University — Master of Science in Education • 2010

Purdue University — Bachelor of Arts • 2006

TECHNICAL SKILLS

AI / LLM: LangChain, OpenAI API, Qwen2-7B, Tool Calling, Prompt Engineering, RAG, FAISS, Vector Databases

ML / Quantitative: XGBoost, Prophet, LSTM, Markov Chains, Statistical Jump Models, Regime Detection, scikit-learn, pandas, NumPy

Data Engineering: PostgreSQL, SQLite, DuckDB, Redis, Parquet, Docker, API Development, Web Scraping, ETL Pipelines

Research / Medical: Survival Analysis, Clinical Data, Statistical Modeling, Peer Review Collaboration, Biostatistics

Computational Biology: Sequence Analysis, Bioinformatics Pipelines, Thermodynamic Modeling, BLAST, Genomic Data Processing

Cloud / Infrastructure: Google Cloud, Async Processing, JSON Schema Validation, Bulk Database Operations