RYAN CLARK

Data Scientist at Harvard Medical School and HHMI
Boston, MA | rpclark219@gmail.com | (718) 938-5299 | www.linkedin.com/in/ryanclark219 | www.ryanpclark.com

EDUCATION & TECHNICAL STACK

Boston College, **Morrissey College of Arts and Sciences** Chestnut Hill, MA
Bachelor of Science in Computer Science, Minor in Finance
August 2020 – May 2024

- **Relevant courses:** Financial Data Analytics – Data Science – Fundamentals of Finance – Corporate Finance – Investments – Financial Accounting – Logic & Computation – Randomness & Computation – Natural Language Processing – Computer Vision – Linear Algebra – Multivariable Calculus – Differential Equations

Programming Languages: Python – SQL – R – Bash – HTML – CSS – JavaScript
Data Engineering & Cloud: AWS (EC2, S3, DynamoDB, RDS, SageMaker, EMR, Glue) – Linux – ETL pipelines – Flask
Databases: DynamoDB – RDS – SQL Server – PostgreSQL – MongoDB
Analytics & Visualization: Tableau – Power BI – Pandas – NumPy – Matplotlib – BeautifulSoup – Selenium

WORK EXPERIENCE

Data Scientist at Harvard Medical School and HHMI (Bioinformatics Domain) Boston, MA
June 2024 – Present

- Designed and deployed scalable data pipelines in Python, R, and Linux to process terabyte-scale structured, unstructured, and semi-structured data, enabling reproducible, high-throughput analysis across 1B+ records.
- Applied statistical modeling and machine learning to identify trends, anomalies, and performance drivers in complex datasets, enabling more informed strategic and operational decisions.
- Built interactive dashboards and automated reporting workflows that translated analytical results into clear, actionable insights for cross-functional teams, improving the speed and clarity of data-driven decisions.

Teaching Assistant at Boston College Chestnut Hill, MA
August 2022 – May 2024

- Delivered technical instruction and mentorship for 100+ students, translating complex analytical and computational concepts into accessible explanations for diverse audiences.
- Provided real-time code reviews and hands-on debugging in Python (NumPy, pandas, scikit-learn, OpenCV), building students' applied proficiency with industry-standard tools for data analysis and modeling.

Data & Systems Intern at Payette Boston, MA
May 2023 – August 2023

- Automated data tracking and reporting workflows using Python, SQL Server, BeautifulSoup, and Selenium to collect, clean, and integrate web-sourced data, improving accuracy and reducing manual processing time by 80%.
- Created diagnostic tools and technical documentation that streamlined data-driven issue resolution, minimizing downtime and enhancing system performance.

PROJECTS

Large-Scale Data Engineering & Pattern Mining (Co-First Author, high-impact journal – in review)
June 2024 – August 2025

- Served as first co-author (under review at *Molecular Cell*, impact factor 16.6) applying advanced data analytics to identify novel sequence patterns and interaction rules in large-scale datasets.
- Built scalable ETL pipelines in Python and R to process billions of records from structured and unstructured sources, enabling high-throughput pattern mining and discovery of predictive features.
- Implemented robust data validation, error handling, and CI/CD integration to maintain data quality, reproducibility, and processing efficiency in high-volume analytics environments.

MarketMotif AI – ML-Driven Stock Market Volatility Pattern Detection Inspired by Genomics
November 2024 – August 2025

- Designed and deployed an end-to-end data engineering and analytics platform for financial time-series, applying advanced pattern mining algorithms to detect and classify volatility regimes.
- Engineered predictive ML features from recurring volatility motifs, driving insights into asset regime shifts and market structure.
- Architected API ingestion, targeted web scraping, and AWS-based infrastructure (IAM, DynamoDB, EC2, Lambda, Route 53, SNS) to enable real-time, fault-tolerant data ingestion, transformation, and model execution at scale.

SignalFrame – Python Library for High-Speed Signal Processing
March 2025 – June 2025

- Created a production-grade Python library for real-time, large-scale data processing, featuring optimized algorithms for interval matching, normalization, and statistical testing; released on PyPI for enterprise and open-source adoption.
- Engineered for institutional-scale performance, delivering 50× faster execution and 2,000× lower memory usage than industry-standard tools, enabling integration into high-frequency, data-intensive analytics pipelines.

CERTIFICATIONS

AWS Machine Learning Specialty – AWS Solutions Architect Associate – AWS AI Practitioner – AWS Cloud Practitioner – Databases and SQL for Data Science with Python – Fundamentals of Databricks – Fundamentals of Visualization with Tableau