

JASON CHARWIN

+1 (980) 680-9658 | jason.charwin360@gmail.com | Chapel Hill, NC, USA | linkedin.com/in/jasoncharwin05/ | jasonindata.com/

EDUCATION

University of North Carolina - Chapel Hill
Bachelor's, Data Science

August 2024 - June 2027
GPA: 3.7

PROFESSIONAL EXPERIENCE

Wells Fargo

Chief Data Office Intern

Remote

May 2026 - Present

- Analyze large-scale enterprise datasets using Python, SQL, and statistical analysis to identify operational trends and support data-driven decision making across internal banking systems.
- Develop data pipelines and SQL-based ETL workflows to ingest, clean, and transform complex business data used for reporting, forecasting, and analytics dashboards.
- Apply regression analysis and statistical modeling techniques to evaluate business performance metrics and identify opportunities for process optimization and operational efficiency improvements.
- Collaborate with cross-functional teams of engineers and analysts to translate analytical insights into production data products, dashboards, and internal analytics tools.

UNC AI Consulting

AI Developer - PORCH Nonprofit AI Chatbot Project

Chapel Hill, NC, USA

February 2026 - Present

- Built an AI-powered chatbot to support PORCH Communities, a national nonprofit combating food insecurity, enabling volunteers and families to quickly access resources and program information.
- Reduced manual support workload for nonprofit staff by implementing an LLM-driven FAQ and navigation assistant, allowing users to retrieve answers and links directly from the organization's knowledge base.
- Designed backend logging and analytics workflows to capture chatbot usage metrics and common user questions, enabling the nonprofit to improve volunteer resources and information accessibility.

UNC School of Data Science and Society

Undergraduate Researcher

Chapel Hill, NC, USA

December 2025 - Present

- Built scalable Python data pipelines and ETL workflows to preprocess high-dimensional genomics and epigenomics datasets, enabling downstream machine learning and statistical modeling experiments across thousands of biological features.
- Implemented and benchmarked deep tensor factorization and network-based representation learning models for disease gene prediction, using Python, PyTorch, and scientific computing libraries to evaluate performance across multiple datasets.
- Designed reproducible ML experimentation workflows using Jupyter, Git, and version-controlled pipelines, improving research reproducibility and enabling faster iteration on feature engineering and model evaluation.
- Applied data processing, dimensionality reduction, and exploratory data analysis techniques to multi-omics biological datasets to support computational biology research and predictive modeling.

PROJECTS & OUTSIDE EXPERIENCE

F1 Corner Analysis - [Link to project](#)

Developer and Data Analyst

Remote

June 2025 - Present

- Built a telemetry analytics system that processes Formula 1 racing data using FastF1, enabling corner-level analysis of braking, throttle, and tire behavior across drivers.
- Applied time-series clustering and lap-delta regression to quantify performance differences between drivers and identify optimal racing lines.
- Developed an LLM-powered natural language interface that converts user prompts into telemetry visualizations, improving accessibility for non-technical users.
- Deployed the system as a Next.js + Python API application, enabling interactive performance analysis of racing telemetry.

Brain Hemorrhage Detection Model

- Developed a deep learning image classification pipeline for detecting intracranial hemorrhage using a ResNet50 convolutional neural network architecture.
- Implemented data preprocessing, augmentation, and model training workflows in PyTorch, enabling reproducible experimentation across medical imaging datasets.
- Evaluated model performance using classification metrics and validation pipelines, supporting reliable clinical prediction experiments.
- Built an interactive real-time inference interface using Streamlit, allowing users to upload medical scans and receive model predictions.

Advanced SQL Lineage & Column Extraction Engine - [Link to project](#)

December 2025 - Present

- Built a Python static analysis system that extracts fully-qualified table.column references from complex SQL queries, enabling automated lineage mapping across large query repositories.
- Improved lineage accuracy by implementing AST-based parsing and scope-aware alias resolution, allowing correct column tracking across nested queries, correlated subqueries, and CTE chains.
- Increased parsing robustness by supporting 30+ SQL edge cases including recursive CTEs, APPLY joins, window functions, PIVOT/UNPIVOT operations, and table-valued functions.
- Enabled multi-dialect compatibility across T-SQL, PostgreSQL, MySQL, Snowflake, and BigQuery, allowing the system to analyze real-world enterprise SQL workloads.

SKILLS

Programming: Python, SQL, Java, TypeScript, R

Machine Learning/AI: Deep Learning, Pytorch, Scikit-learn, Feature Engineering, Machine Learning, 3D Modeling, LLM Applications, LLM, Prompt Engineering, Natural Language Interfaces

Data Science/Data Engineering: Pandas, NumPy, Statistical Inference, Regression Analysis, Time-Series Analysis, Data Visualization, Matplotlib, ETL Pipelines, Data Analysis, Data Cleaning, Data Modeling, SQL Query Optimization

Web/Application Development: React.js, Next.js, Flask, Node.js, REST APIs, Tailwind CSS

Infrastructure/Tools: Git, DevOps, GitHub Actions, MongoDB, Vercel

Languages: Spanish, Tamil