

ALEX LIN, MS

DATA SCIENTIST

919-675-5007 | alexlindatascience@gmail.com | [GitHub](#) | [LinkedIn](#) | [Website](#)

Performance-driven data scientist delivering projects in the areas of Big Data (Machine Learning, Data Science), executing data and research analysis, and communicating findings to technical and non-technical audiences. Exceptional analytical and problem-solving skills, able to identify KPIs and discover non-obvious patterns, behaviors, and dependencies in data. Effective communication, data presentation, and visualization skills. Nimble and able to adapt and change in a challenging environment, as part of a team or individually. Conversational in Mandarin.

- ▶ As a data scientist intern at Raleigh Stormwater, developed a time series analysis for the Walnut Creek floodplain to develop predictive models that improved evacuation lead times by 30 minutes.
- ▶ Built targeted user segments by analyzing 600+ GB of behavioral data, identifying high-engagement users, and optimizing Axios' content strategy.
- ▶ Created an interactive Tableau dashboard for the NC State Women's Tennis coach, increasing staff engagement by 40% and enabling data-driven coaching decisions.
- ▶ Developed a Retrieval-Augmented Generation (RAG) movie recommender app with an AI-Powered Personalization & Recommendation Systems.

Competencies –

Data Visualization & Dashboard Development • Machine Learning Algorithms • AI Design • Cloud Data Engineering • Data Ethics & Compliance • Computer Vision • Quantitative & Qualitative Research Methods • Data Research & Analysis • API Integration • LLM Deployment • Complex Problem-Solving Skills • Team Building • Communicating to Non-Technical Audiences

EDUCATION

Master of Science (MS), Analytics, Institute for Advanced Analytics, GPA: 3.7
North Carolina State University, Raleigh, NC – 4/2025

Bachelor of Science (BS), Statistics and Analytics, GPA: 3.7
University of North Carolina at Chapel Hill, Chapel Hill, NC – 5/2024

Capstone – Machine Learning Project: “Acoustic Features of Songs Mood Classification”

PRACTICUM

Axios | Communication Lead | Remote

8/2024 to 4/2025

- Faced with 600+ GB of behavioral data that exceeded Jupyter Notebook's memory limits, requiring analysis to identify high-engagement users for newsletters and ads.
 - Migrated data to PostgreSQL, designed and implemented an ETL pipeline to consolidate, clean, and batch-load 600+ GB of behavioral data into a PostgreSQL database, significantly reducing load times and enabling scalable, reliable analysis for downstream modeling and reporting saving the team over 40 hours per month on manual data handling.
 - Reduced full dataset processing time from 8 hours to 45 minutes, enabled stable analysis of terabyte-scale behavioral patterns through SQL window functions.
- Axios needed a tool to segment users and track key engagement metrics across different audience groups.
 - Developed an interactive dashboard that visualized audience trends and segmented users based on behavioral data. Incorporated engagement metrics for each segment, enabling users to compare engagement, retention, and content interaction across different audience types.
 - Introduced mapping to allow dashboard users visibility of location of most engaged users. Focused on clear, actionable data storytelling to make insights accessible for technical and non-technical stakeholders.
 - Empowered the Axios team to quickly identify high-value user segments and monitor the effectiveness of targeted content strategies, leading to more data-driven decision-making and improved audience engagement.
 - Dashboard's intuitive design and tailored metrics received positive feedback for enhancing visibility into user trends and supporting ongoing optimization efforts.

TECHNICAL SNAPSHOT

Programming Languages: Python, SQL, R, Git
Libraries: Pandas, NumPy, SciPy, Matplotlib, Seaborn, TensorFlow, PyTorch, Keras, OpenCV

Cloud Technologies: AWS, Snowflake, MongoDB

Data Visualization: Tableau, Plotly Dash, Streamlit, Power BI

Certifications: PCEP—Certified Entry-Level Python Programmer

PROFESSIONAL EXPERIENCE

Raleigh Stormwater | Data Scientist Intern | Raleigh, NC

6/2023 to 8/2023

- Designed the Walnut Creek floodplain prediction model project scope. Analyzed water levels and developed an indicator identifying the accelerated rate increases at one time.
 - Developed an automated process that updates a Power BI dashboard with new water level data daily. Used time series analysis and ARIMA models to predict timing of events and identify patterns.
 - Integrated GIS mappings for interactive visualizations, which combined with the dashboard, as a method to communicate findings to the Stormwater Advisory Commission.
 - Improved evacuation lead times by 30 minutes, shaped mitigation strategies, and influenced city policy through presenting findings.
- Tasked with analyzing the Raleigh Stormwater contractor project budget and visualizing funds allocated and spent across distinct phases of each project, which valued over \$10s of millions of dollars in contracts.
 - Consolidated and cleaned budget data and created clear visualizations using Excel, Matplotlib, and Seaborn to illustrate spending trends and project cycle expenditures for non-technical stakeholders.
 - Delivered intuitive figures that improved stakeholders' understanding of budget allocation over project lifecycles, supporting more informed project planning, and earning positive feedback from the team.

Additional Relevant Projects –

NC State Women's Tennis Analytics Dashboard (Client-Facing Project)

- Built the coaching staff a tool that enabled data visualization for success metrics for each player. Developed a Tableau dashboard visualizing KPIs for 50+ players over 200+ matches, optimizing UI/UX for usability.
 - Increased dashboard engagement by 40% compared to previous dashboard solution and improved data-driven decision-making for coaching staff.

RAG Movie Recommender (Personal Project)

Designed a Retrieval-Augmented Generation (RAG) movie recommender web app using Streamlit, integrating the TMDB API to access up-to-date film and TV metadata and deploying the DeepSeek R1 large language model locally with Ollama.

- Developed an ETL pipeline to extract user preferences, transform and filter recommendations by genre, year, and rating, and load enriched results into the app.
- Leveraged RAG to generate concise, AI-powered explanations for each recommendation, making the logic behind suggestions transparent and user-friendly.
- Delivered an interactive application that adapts to individual user inputs, provides tailored recommendations, and explains each suggestion in plain language.

Mario Party DS Automation (Personal Project)

- Researched and implemented Python-based automation solutions for multiple minigames.
 - Used OpenCV and NumPy for real-time image processing and object detection, and PyAutoGUI for simulating mouse and keyboard inputs.
 - Developed custom algorithms for each game, including RGB-based prompt detection and binary masking for object identification. Iterative refinement of automation logic reduced false positives by 40% and improved task completion speed by 25%, resulting in 98% overall automation accuracy. Managed emulator focus and capture for seamless operation.
 - Successfully automated gameplay for several minigames, achieving high accuracy and responsiveness.