

ANDREW LEE

+1 (757) 940-8688 | lee.andrew.jooyoung@gmail.com | Boston, MA | leeanddrew.github.io

EDUCATION

Harvard University

Master of Science in Data Science

Cambridge, MA

Dec. 2026 (Expected)

Courseworks: Natural Language Processing, MLOps, Linear and Generalized Linear Models, Optimization Methods

William & Mary

Bachelor of Science in Mathematics; GPA: 3.98/4.00; Phi Beta Kappa

Williamsburg, VA

May 2025

Courseworks: Probability Theory, Mathematical Statistics, Statistical Learning, Regression Analysis, Data Structures

PROFESSIONAL EXPERIENCE

Data Scientist (Part-Time)

Aug. 2025 – Present

Data Science Intern

June 2025 – Aug. 2025

E. & J. Gallo Winery

Modesto, CA

- Designed and deployed a Neo4j graph database with a fully automated ETL pipeline, integrating AWS Lambda for daily updates and schema versioning in a production-ready environment
- Built and trained a SARIMAX time series model to forecast monthly wine production with 6% error, enabling the operations team to make data-driven decisions using prediction intervals
- Simulated tank utilization using tank cycle time as a key efficiency metric, leading to operational insights of increased wine throughput from 9M to 22M gallons
- Accelerated the team's MLOps initiative through containerization, CI/CD pipelines, data versioning, and DevOps best practices for deploying multiple ML models into production

Data Analyst Intern

May 2022 – Aug. 2022

Council on International Educational Exchange (CIEE)

Seoul, South Korea

- Designed 3 study abroad programs with tailored themes, improving inquiry-to-application conversion by 20%
- Built and pitched Tableau dashboards that accelerated itinerary approvals by 2 weeks and resolved data inconsistencies across 200+ students' flights, housing, and survey information in Excel
- Ran A/B tests on post-event survey data to evaluate the cost-efficiency of guest speaker formats, influencing a budget pivot and saving \$10,000 per program cycle

RESEARCH EXPERIENCE

Enhancing Sandhill Crane Counts Using Deep Learning – with U.S. Fish and Wildlife Service [[Code](#)]

- Developed multi-class object detection model for counting migratory birds on 47,000+ thermal images
- Reduced operational costs by an estimated \$20,000/year by enabling weeks-long surveys to be completed in hours
- Deployed a user-friendly Dockerized app on AWS EC2 for wildlife researchers to easily run and evaluate the model
- Presented research findings to non-technical stakeholders including the CEO of Conservation International

Spatiotemporal Analysis of Human Response to Wildfire – with Harvard Medical School [[Code](#)] [[Paper](#)]

- Implemented a Bayesian spatiotemporal distributed lag model to forecast public response to wildfire smoke PM2.5
- Conducted exploratory data analysis (EDA) in R and leveraged Selenium to automate data collection
- Identified increased sensitivity to smoke PM2.5 among individuals experiencing poverty with a 22% surge in air pollution searches and an 18% rise in air purifier interest

Machine Learning in Hypersonic Vehicles – with HyperStats Lab at William & Mary [[Code](#)]

- Built and cross-validated supervised learning models (Linear Regression, Random Forest, SVM) to predict Shock-Train Leading Edge (STLE) in hypersonic vehicles using high-dimensional sparse pressure sensor data
- Improved R-squared of 0.64 from a previous study to 0.89 by engineering time-series features
- Trained a Recurrent Neural Network (RNN) and achieved successful generalization to 900K+ test observations

TECHNICAL SKILLS

Programming Languages: Python (Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn), SQL, R, Bash/Shell

MLOps/DevOps Tools: AWS (S3, EC2, Lambda, SageMaker, Redshift, QuickSight), Docker, DVC, Git, CI/CD

Certifications: AWS Certified Cloud Practitioner, Microsoft Office Specialist