Calvin Carter

505-589-8388 ◊ calv2n@berkeley.edu ◊ github.com/calv2n ◊ linkedin.com/in/calvinmcarter
Berkeley, CA

EDUCATION

University of California, Berkeley *Exp. Dec 2025*
Master of Arts in Statistics

Relevant Coursework: Advanced Probability, Advanced Statistics, Statistical Computing.

University of California, Berkeley *May 2024*
Bachelor of Arts in Data Science with emphasis in Applied Math & Modeling GPA: 3.63

Relevant Coursework: Data Engineering; Principles & Techniques of Data Science; Data, Inference, & Decisions; Data Structures; Concepts in Computing with Data; Linear Programming & Network Flows; Linear Algebra.

Certification: Mastering Data Analysis in Excel, Duke University

SKILLS

Technical: Python, SQL, R, MongoDB, Java, Excel, Machine Learning, Statistical Modeling, Linear Algebra, Linear Programming, Statistics, Causal Inference, Data Visualization, Computer Science, Git/GitHub, NumPy, scikit-learn.

Soft Skills: Ability to Effectively Communicate, Decision Making, Adaptability, Curiosity, Teaching, Critical Thinking, Attention to Detail, Problem Solving Skills, Focus, Leadership, Presentation, Ambition, Creativity, Collaboration.

EXPERIENCE

Data Science Projects

Gradebook - *R, Plotly, Shiny, Git/Github* *Jan 2024 - Jun 2024*
- Collaborated with Berkeley statistics professor and peers to RShiny web application for instructor grades.
- Designed and implemented intuitive dashboard that provides insightful statistics and visuals for instructors.

Email Spam Detection - *Python, NumPy, scikit-learn, SVM, Machine Learning* *Feb 2024*
- Utilized Support Vector Machine (SVM) model to classify emails as spam or non-spam.
- Implemented k-fold cross-validation for robust model evaluation and achieved a final test accuracy of 86%.

Diabetes Risk Classification - *Python, NumPy, Random Forest, Logistic Regression, scikit-learn* *Dec 2023*
- Applied Exploratory Data Analysis to develop Logistic Regression and Random Forest classification models on diabetes risk. One for higher interpretability and the other for high accuracy of 98% using Scikit-learn.

Predicting Housing Prices - *Python, NumPy, Linear Regression, scikit-learn, Machine Learning* *Oct 2023*
- Created multiple linear regression model to predict housing prices in Cook County, Illinois.
- Applied EDA with data visualization and feature engineering techniques and to minimize model RMSE.

UC Berkeley Statistics Department

Graduate Student Instructor - Concepts in Computing with Data *Aug 2024 - Present*
- Deliver engaging lab sections for a cohort of 76 students, covering foundational and advanced statistical computing concepts in R.
- Guide students in adopting best practices for data analysis and statistical computing, including effective use of R packages, reproducible research techniques, and data visualization principles.

Co-Founder & Facilitator - Intro to Data Visualization DeCal Course *Jun 2023 - July 2024*
- Developed syllabus and 14 weeks of course content including lectures, labs, homework, and video tutorials.
- Conducted surveys to adjust and improve course material throughout each semester. Saw an increase of over 300% in enrolled students between Fall and Spring semesters as a result of improvements.