David Lee

davidlee092802@gmail.com | Brea, CA | linkedin.com/in/david-lee-305b9422b/ | github.com/davvidlee

Education

Pomona College – B.A in Computer Science   May 2025

Skills

Languages: Python, SQL, JavaScript, Java, Typescript, Haskell, C#
Libraries/Frameworks: Pandas, NumPy, scikit-learn, PyTorch, Altair, React
Tools: Excel, VS Code, JIRA, Git, Firebase, AWS, DynamoDB, Microsoft Office
Database: PostgreSQL, SQLx, Firebase
Testing: Playwright, assert_cmd, cargo

Experience

QA Support/QA Engineer, Pomona College ITS – Claremont, CA   Feb 2023 – Dec 2024
- Developed automated testing scripts with Playwright in Python, cutting manual QA time by 2+ hours.
- Cleaned and validated Excel datasets, resolving 30+ discrepancies in academic records.
- Conducted user experience analysis, driving website improvements using JavaScript and HTML.
- Managed data-driven bug tracking via JIRA, resolving 50+ issues using structured metrics.

Research Assistant Programmer (Backend Integration), FAIM Lab – Claremont, CA   Feb 2025 – Present
- Built and tested robust data pipelines integrating HTTP APIs with PostgreSQL databases.
- Wrote integration tests for Rust-based backend, verifying authentication and access patterns.
- Used SQL and CLI-based workflows to automate database setup and simulate test scenarios.

Computer Science TA and Mentor, Pomona College – Claremont, CA   Aug 2023 – Dec 2023
- Mentored students in functional programming, guiding them through key concepts and debugging assignments.
- Conducted two weekly office hours, providing targeted assistance on coursework and programming challenges.
- Collaborated with the professor by identifying student difficulties, offering insights to improve instruction, and assisting with grading.

Projects

Baseball Analytics | Python, scikit-learn, NMF
- Designed and implemented a Non-Negative Matrix Factorization (NMF) model for analyzing baseball analytics data using scikit-learn and PyTorch.
- Analyzed 345,000+ at-bats from April to June 2024, covering 559 unique batters and 689 unique pitchers.
- Utilized our NMF model to perform K-Means clustering, uncovering that contact hitters have an average bat speed of 65.6mph and power hitters have an average bat speed of 71.7mph.

Obesity & Food Affordability Analysis | Python, Pandas, Altair, BeautifulSoup
- Analyzed food prices, GDP (PPP), and obesity rates across 90+ countries to investigate affordability's role in public health.
- Cleaned and merged data from Numbeo and Wikipedia; built visualizations showing fast food affordability strongly correlates with obesity rates.
- Used Python, Pandas, and Altair to create clear visual storytelling with public datasets.

Climate & Demographic Mapping | Python, Geopy, CensusGeocode, Plotly
- Mapped U.S. locations to Census data to visualize how climate change and pollution intersect with income, education, and population.
- Used geocoding and the Census API to enrich locations with demographic features and visualize trends with Plotly and Seaborn.
- Enabled analysis of environmental disparities across neighborhoods.