Andrew Nguyen

(408) 946-4400 | andrew.nguyen@gmail.com | andrwngyn.github.io | linkedin.com/in/andrw-ngyn

EDUCATION

University of California, Irvine (UCI) | Master of Data Science   Expected Dec. 2025
- Probability and Statistics, Databases and Data Management, Artificial Intelligence

University of California, Santa Barbara (UCSB) | B.S. Statistics and Data Science   Dec. 2023
- Stochastic Processes, Statistical Machine Learning, Time Series, Bayesian Data Analysis

SKILLS

Python | R | SQL | Bash | Git/Github | TensorFlow/Keras | seaborn | Flask | ggplot | tidymodels | tseries

EXPERIENCE

Engineering Intern

Picarro, Inc. | Santa Clara, California   Apr. 2024 - Aug. 2024
- Spearheaded development and deployment of a customized gas leak investigation tool using GeoPandas, tailored for various clients, and presented results to executives.
- Implemented data pipelines for an internal Slack bot, including EU server integration, improving data handling efficiency and responsiveness.
- Automated multiple scripts for internal data processing, integrating with Slack, Box, and Geoserver APIs to boost efficiency and reduce manual effort.

Data Analytics Intern

Picarro, Inc. | Santa Clara, California   Jun. 2023 - Sep. 2023
- Developed neural network model for detecting instrument measurement anomalies with 93% F1 score and 0.01% class imbalance. Model automated using a Slack chatbot on AWS EC2 to notify Field Application Engineers, eliminating 100% of manual labor. Created pipelines to auto-feed data into predictive models
- Extracted performance data from Jira via PostgreSQL into Apache Superset for a performance metrics dashboard
- Executed MSSQL queries for computational diagnostics and data provision to engineers

PROJECTS

Medical Study Service Area (MSSA) Scoring Algorithm   Aug. 2024

Geospatial analysis of MSSAs to prioritize medical resource allocation with GeoPandas and Folio
- Quantified Medical Study Service Areas (MSSAs) by their access and quality to healthcare facilities
- Expanded an existing binary scoring algorithm to a nuanced ranking system for MSSAs, including additional metrics for prioritization.
- Designated areas of high priority for medical resource allocation

US Electricity Production Forecast   Mar. 2023

Time Series Analysis of US Electricity Production Types in R
- Forecast and analysis of US Electricity Production types using time series models in R. Production values for various production types (fossil fuels and renewable energy) are forecasted using SARIMA models
- Independent research of US Electricity Production type and behavior, and gained insight into energy production trajectory for the future