

STEVEN R WILSON

Palmdale, CA | 562.822.3002 | stevenwilson.data@gmail.com
[stevenwilson-data.github.io](https://github.com/stevenwilson-data) | github.com/stevenwilson-data | [linkedin.com/in/stevenwilson-data](https://www.linkedin.com/in/stevenwilson-data)

TECHNICAL SKILLS

Languages & tools: Python (pandas, scikit-learn, Plotly), R (glmnet, time series), SQL, Excel, Git / GitHub

Methods: time-series forecasting, multiple regression, LASSO and model selection, k-means clustering, bootstrap and cross-validation, survival analysis (Kaplan–Meier), anomaly detection, data-quality auditing

INDEPENDENT DATA ANALYSIS PROJECTS | 2026

When Traffic Breaks Down · Caltrans PeMS

Python

- Scripted restart-safe collection of 120 daily Caltrans PeMS files (~2.2 GB); audited detector coverage and restricted analysis to fully observed intervals.
- Compared three breakdown definitions that found 1,457, 10 and 7 events in the same four months; tested results with matched controls, leave-one-out checks and Kaplan–Meier recovery analysis.
- Located a recurring bottleneck: 172 persistent queue runs between two adjacent stations vs. 36 for the next-highest pair.

Ames Housing Regression · R / Python replication

R + Python

- Replicated a full regression workflow in R and Python on identical train/test splits; selections were perfectly nested and test RMSE agreed within \$53 on ~\$21,000.
- Reduced a 221-column design matrix with LASSO and BIC selection; compared candidate models with a 10,000-resample paired bootstrap.

Electricity Demand Forecasting · EIA-930, California

R

- Compared seasonal-naïve, Fourier + ARIMA, 50/50 ensemble and published EIA forecasts on 11 years of hourly demand (97,344 observations).
- Ensemble cut MAE 11% vs. the seasonal-naïve baseline across 8,544 forecasts; traced 84% of the EIA forecast's net underforecast to nine daytime hours.

Customer Lifecycle Segmentation · UCI Online Retail II

Python

- Cleaned ~1.07 million transactions and reconciled 1,210 same-day purchase-cancellation pairs before building account-level features.
- Segmented 5,823 accounts with k-means, validated with silhouette, Davies–Bouldin and bootstrap ARI (0.981); 18% of accounts produced 72.8% of net revenue.

SQL Business Metrics · AdventureWorks

SQL + Python

- Queried three years of AdventureWorks sales in SQL with Python / Plotly reporting; online orders were 87.9% of orders but only 26.7% of merchandise revenue.

PROFESSIONAL EXPERIENCE

NASA Armstrong Flight Research Center / SOFIA · Cal Poly STAR Program

Summer 2013

- Started with no assigned project; identified and scoped a Telescope Assembly sensor study by asking SOFIA personnel what needed analysis.
- Wrote all analysis code in R and compared system behavior across 14 flights, extending earlier per-flight analysis to separate brief spikes from sustained Power Bus anomalies.
- Determined a threshold for warning software to alert operators to power-load problems; applied the same method to gyroscopic drift.
- First author of the resulting research poster; wrote the initial draft and all figures, with review from NASA, DSI and USRA collaborators.

iLEAD SCVi, Santa Clarita · Math Teacher, Grades 8–11

Aug–Dec 2025

Endeavour Middle School, Lancaster · 7th Grade Math Teacher

Jun 2024–Jun 2025

Antelope Valley High School, Lancaster · High School Math Teacher

Aug 2023–Jun 2024

Orange County Dept. of Education, Juvenile Hall · Substitute Teacher

2022–Aug 2023

Scoot Education, Los Angeles · Substitute / SpEd Paraeducator

2022–2023

WyzAnt · Independent Tutor, Mathematics & Statistics

2009–Present

EDUCATION

California State University, Fullerton · B.A., Mathematics

2013