

Sarah Lew

Walnut, CA | sarahl02116@gmail.com | (909) 492-3962 | [LinkedIn](#) | Portfolio: <https://github.com/sarahllew>

EDUCATION

University of California, Santa Barbara

Santa Barbara, CA

Statistics and Data Science, Bachelor of Science (B.S.)

GPA: 3.82 | Expected Graduation: March 2025

Relevant Coursework: Data Science Principles with R and SQL, Transitions into Data Science and Higher Mathematics, Statistical Machine Learning, Design of Experiments, Python Data Structures, Time Series, Regression Analysis, SAS Base Programming

EXPERIENCE

Data Analytics Intern

Redondo Beach, CA

Northrop Grumman Space Systems

June 2024–August 2024

- Developed internal search engine tool with Django frontend for team to identify peers with specific skills and projects
- Transformed SQL queries to expedite ETL process by extracting data from SAP timekeeping and staffing systems
- Authored pre-processing functions for NLP model and improved result accuracy by stemming and tokenizing search texts
- Created 6+ interactive Tableau dashboards to reduce operational costs by tracking power consumption of campus buildings
- Leveraged Agile methodologies to manage and lead sprints for full-stack tool that organizes SAP data for engineering teams
- Designed dashboards for stakeholders to monitor staffing metrics, enabling efficient tracking of job fulfilment timelines

Business Management Contracts and Pricing Intern

San Diego, CA

Northrop Grumman Mission Systems

June 2023–August 2023

- Analyzed 30+ Basis of Estimates to price service packages, ensuring data validation and streamlining team pricing strategies
- Devised Excel statistical formulas and data visualizations, comparing cost data across 8 years to determine repricing needs
- Conducted Rough Order of Magnitude estimate by pricing inputs and creating visibility files for accurate cost projection
- Performed 5+ contract close-outs and delinquencies reports using SAP software, optimizing organization procedures

Recording Analyst Intern

Walnut, CA

Progreen Services, LLC

January 2019–June 2023

- Evaluated yearly and monthly data records and costs from comprehensive reports, monitoring cost trends
- Established 10+ in-depth assessments of invoices to ensure records are precisely compiled with supporting documents

PROJECTS

Predictive Analytics: Sephora Website Product Ratings | R

- Prepared dataset of Sephora website products, processing and tidying data for analysis with dplyr and tidyverse
- Performed visual exploratory data analysis, analyzing potential underlying trends and correlation of variables
- Trained 6 machine learning models including linear regression, k-nearest neighbor, elastic net, and tree-based methods
- Collected RMSE metrics for each model after tuning hyperparameters and visualized best model performance on testing data

Forecasting U.S. Unemployment Rate Post Pandemic | R

- Analyzed stationarity and seasonal components of U.S. Unemployment rate data from 1948 to 2023 with diagnostic plots
- Applied time series analysis methods such as SARIMA and spectral density analysis with periodograms
- Identified seasonal and stationary trends and tested best performing model, accurately forecasting yearly unemployment rate

Chihuahua vs Muffin Classification | Python

- Trained CNN model to classify whether a user's image is more similar to a chihuahua or muffin using TensorFlow and Keras
- Incorporated model into Flask web app with OpenCV, allowing users to upload an image of themselves for classification

Used Car Lot Collection | Python

- Designed verification methods of car objects based on the make, model, year, and price of the car in sorted order
- Implemented program that manages all objects of car inventory and determines the total price and best option for customer

SKILLS

Languages: R | Python | SQL | SAS | HTML | JavaScript

Tools: RStudio | MS SQL Server | VS Code | Tableau | Git | Bitbucket | Jira | Confluence | Azure DevOps | Django | SAP ERP

Libraries: NLTK | Pandas | Scikit-learn | NumPy | TensorFlow | Keras | tidyverse | dplyr | XGboost | ggplot2 | Shiny | knitr | lubridate