

Calvin Raab

Data Scientist

(858) 335 - 9010 | calvin.raab@gmail.com | [linkedin.com/in/calvinraab/](https://www.linkedin.com/in/calvinraab/) | github.com/calvinraab

EXPERIENCE

SAN DIEGO ASSOCIATION OF GOVERNMENTS

San Diego, CA

Data Scientist II

SEPTEMBER 2021 - Present

- Lead Python development for the Quality Assurance team, building end-to-end automated testing pipelines
- Analyze regional data to inform operational, procedural, and policy decision-making
- Serve as the primary code reviewer for the data science department, ensuring code quality and adherence to best practices
- Co-lead the Python User Group and mentor three data scientists
- Contribute to the team moving SANDAG to Microsoft Azure
- Promoted ahead of schedule due to exceptional performance and impact

EDUCATION

UNIVERSITY OF MICHIGAN

Ann Arbor, MI (remote)

Master's in Applied Data Science

MAY 2022-Present

UNIVERSITY OF CALIFORNIA

Los Angeles, CA

Bachelor of Arts, Economics

Graduated: June 2021

PROJECTS

Forecast Quality Control Automation | Python, SQL Server, Power BI, Git

MAY 2023

- Developed an automated system for the compilation and processing of SANDAG's Forecast Estimates saving the QA team one week of work every forecast run
- Developed statistical checks that check the data and raise flags where necessary
- Built visualizations in Power BI to allow SANDAG analysts to discover more insights into Forecast Estimates
- Built a fully functioning user interface for other data scientists to run this application

SANDAG Unemployment Storymap Automation | Python, SQL Server, Socrata API, Git

AUGUST 2022

- Automated end-to-end SANDAG's unemployment data processing pipeline
- Built testing functionality to test the outputs of my automation
- Setup automatic testing using Github Actions
- Saved 7 hours per month for each data scientist on the project

TECHNICAL SKILLS

Languages: Python, R, Microsoft SQL Server, PostgreSQL, Shell/Bash, HTML/CSS

Data Science Tools: Microsoft Azure, Hadoop, Apache Spark, Power BI, Git, ArcGIS Pro, Excel

Libraries/Frameworks: Pandas, NumPy, PySpark, Populationsim, Scikit-learn, Django, unittest, Bootstrap

Subject Matter Knowledge: Statistics, Calculus, Linear Algebra, Cloud Computing, Economics, Demography, Transportation Modeling, Population Forecasting, Finance, Data Science