

Work Experience

Data Scientist & Engineer | [Joby Aviation](#) Jul. 2021 - Present

- Proposed and developed foundational tools used by a data science team of 17
 - * CI build system **4 to 20x faster** than the previous iteration while implementing several new checks which eliminated entire classes of bugs
 - * Distributed framework on top of Spark automating the execution of analyses upon availability of new flight data in **~2,100 LoC**, replacing a series of frequently-failing systems which required maintaining **~30,000 LoC**
- Designed and implemented data strategy for aircraft computational modeling team of 30
 - * Physics-driven data augmentation which **doubled** the amount of data usable for model FAA certification
 - * ETL pipeline in Spark and Numpy standardizing output formats across multiple computational fluid dynamics models, enabling model regression testing and one-to-one comparison between models across **~200** physical metrics used for model certification
- Collaborated with multiple internal teams to build end-to-end data tools acting as force multiplier on their efficiency
 - * Pipeline and dashboard using Scipy, Pandas and Dash automating assessment of aircraft powertrain tests in manufacturing, reducing time to results from **2 hours to 30 minutes**

Data Scientist Intern | [Fullpower Technologies](#) Jan. 2021 - Jun. 2021

- Ported Tensorflow models to AWS Inferentia chips, **doubling** inference throughput at fixed cost
- Ran comparative analysis of sleep studies against medical device specifications, which was used in the FDA approval filing for this device

Research Assistant | [UC Santa Cruz Genomics Institute](#) Sep. 2020 - Jun. 2021

- Derived novel EM solver for genotype generation, with [implementation in numba](#) **100x faster** than equivalent Python code.
- Investigated generative models for genotypes, finding that state-of-the-art GAN architectures have trouble scaling to large gene sequences

Research Intern | [Center for Research in Open Source Software](#) Jul. 2020 - Sep. 2020

- Added the ability to run Jupyter through the Docker and Singularity backends of the [Popper](#) task automation engine
- Developed automation workflows for computational research enabling end-to-end replication of a research project

Web Developer (contractor) | [Merit Educational Consultants](#) Jul. 2020 - Oct. 2020

- Built supply management app with React and Firebase, which was used for a year without requiring any updates or bug fixes

Data Scientist Intern | [Startup Genome](#) Apr. 2020 - Jun. 2020

- Built analysis pipeline in Pandas for global survey data which reduced time to view new survey results from **30 to 2 minutes**

Technical Skills

Programming languages: Python, SQL, Scala, Javascript, CSS

Frameworks & libraries: Apache Spark, SciPy, Pandas, FastAPI, Plotly, PyTorch, React, LiHaoyi Scala

Tools & platforms: Docker, Databricks, Kubernetes, Grafana, Make, Terraform, PostgreSQL, AWS (S3, ECR, SQS)

Education

University of California, Santa Cruz Aug. 2017 – May 2021

B.A. Mathematics (Honors), B.A. Computer Science (Honors)

Coursework towards and M.A. in Mathematics

Volunteer Experience

President | [Data Science @ UCSC](#) Jan. 2019 - Jan. 2021

- Led student team to **top 5% global** scores in several machine learning competitions on Kaggle and Driven Data
- Placed **1st at CruzHacks 2020** with a data visualization website aggregating data from Facebook and Google to bring transparency to political ad spending in the 2020 election