

PROGRAMMING
DATABASES & WEB FRAMEWORKS
CLOUD COMPUTING & DASHBOARDING
TOOLS & LIBRARIES
MACHINE LEARNING

Python, SQL, Datalog
Postgresql, Datomic, MongoDB, Flask, Django
AWS, EC2, S3, Sagemaker, Redshift, Athena, Tableau, Power BI
Pandas, Numpy, Matplotlib, Plotly, Sklearn, Statsmodels, BeautifulSoup, Selenium
Regression, Classification, Clustering, Time Series Analysis, Natural Language Processing

EXPERIENCE

THERMO FISHER SCIENTIFIC

Research and Development Data Scientist

South San Francisco

Feb. 2021 to Current

- Built a **scalable, end-to-end ETL process** to achieve **real-time monitoring** of SARS-CoV-2 RT-PCR test data from labs and universities using **Amazon AWS** and **Power BI**. Listed as third author on a **symposium abstract** based on this work
- **Lead the development** of the **data pipeline**. Heavily **refactored** and **optimized** the existing data pipeline for **scalability** through clever use of **parallelization** and **batch processing**, cutting the time it takes to transform 2000 csv files from 2 hours to 3 minutes. Add new **KPIs** and support **backward compatibility**
- **Analyzed** performance and results of sample retests using **Python** and **SQL**, enabling stakeholders to make data-driven business-decisions
- Created new **Plotly interactive visualization** for 96- and 384-well plates, enabling the Data Science team to rapidly **identify anomalies** in new data

INVITAE

Bioinformatics Data Scientist

San Francisco

Jan. 2020 to Sept. 2020

- Discovered the **root cause** for a high-visibility, time-sensitive issue on a new assay that resulted in batch failures impacting 500+ samples
- Analyzed **trends** using **Python** and **SQL** to **monitor next-generation sequencing (NGS)** production-line data for anomalies
- Efficiently diagnosed **customer issues** by writing a **Python script** to **automate** generation of standardized **visualizations** and **statistics**
- Wrote maintainable **Datalog** and **SQL** queries for **Flask** app used to migrate data from **Datomic** to **Amazon Redshift** for consumption by **Tableau**
- Built an 8-page **Tableau dashboard** to enable our operations team to carefully monitor a new product launch

NEXTBEE MEDIA

Data Scientist

San Mateo

Sept. 2019 to Dec. 2019

Led **development** of the Lighthouse App from inception to deployment. Available at: <https://lighthouse.nextbee.com/>

- Designed the **relational database model** and wrote the **Python** scripts used to build customer profiles based on orders data
- **Segmented** customers into **tier groups** based on features identified through **ecommerce domain knowledge**
- Used **logistic regression** and **random forest** to predict the likelihood of each customer making another purchase based on their purchase history
- Used **time-series forecasting** to predict future revenue and number of new customers

BIOVERATIV, FORMERLY TRUE NORTH THERAPEUTICS

Research Associate 2

South San Francisco

Jan. 2017 to Mar. 2019

- Led **protein-engineering project** on of lead antibody drug (BIVV009). Handled all **cloning and expression** of antibody variants and most of protein purification for the lab. **Collected binding and efficacy (KD and IC50) data** and performed **regression analysis**, revealing a log-linear relationship between KD and IC50. This enabled us to select which variants to use in downstream experiments.
- Led **research project** to make a protein complex of BIVV009 and its target (C1s of the complement immune system) for crystallography study.
- Wrote **Python script** to automate design of short DNA oligos, reducing a task that normally took 2 hours/week of manual labor to a 3 minute script with some minor follow-up.
- Wet lab: tissue culture, protein expression and purification, ELISAs, affinity and kinetics characterization, crystallography

GENE YEO LAB, UCSD

Staff Research Associate 1

La Jolla

May 2013 to Nov. 2016

- Developed **standardized protocols** to create tagged cell lines for **next-generation sequencing** (ENCODE)
- **Co-authored a Neuron paper** that included my experiments on investigating the mechanism for how a mutation in an RNA-binding protein (hnRNPA2B1) can cause ALS
- **Co-authored a Cell paper** that included my experiments on using a new genome-editing technology (CRISPR/Cas9) to track RNA in live cells
- Wet lab: **fluorescent cell imaging**, **next-generation sequencing**, high-throughput sequencing, CRISPR-Cas9, qPCR

FEATURED DATA PROJECTS

CLIMBING LOG WEB APPLICATION

Live app at: <https://harrisonized-climbing-app.herokuapp.com>

Blog post at: <https://harrisonized.github.io/2020/11/05/climbing-dashboard.html>

Built a **Flask app** to function as a **live dashboard** that keeps track of my climbing records. On the back-end, executes **SQL queries** to extract data from **Heroku Postgres**, transform the data, and generate **Plotly interactive visualizations**. On the front-end, displays the **visualizations** in logically organized web pages. Visualizations are **cached** to streamline data access and improve user experience. Users can **upload custom data** to generate their own figures.

EDUCATION

Metis Data Science Bootcamp

Apr. 2019 to June 2019

University of California, San Diego

2015

Double Major: B.S. Physics, B.S. Physiology & Neuroscience