HARRISON WANG
BIOINFORMATICS DATA SCIENTIST

harrison.c.wang@gmail.com
https://harrisonized.github.io/
https://www.linkedin.com/in/harrisonized/
https://github.com/harrisonized

PROGRAMMING: Python, SQL, Datalog
INFRASTRUCTURES & DATABASES: Postgresql, Datomic, MongoDB, AWS
MACHINE LEARNING: Regression, Classification, Natural Language Processing, Clustering, Time Series Analysis
TOOLS & LIBRARIES: Tableau, Plotly, Pandas, Numpy, Matplotlib, Sci-kit Learn, StatsModels, BeautifulSoup, Selenium, Samtools, Bedtools
WEB FRAMEWORKS AND DOMAIN KNOWLEDGE: Flask, NGS

EXPERIENCE

INVITAE
San Francisco
Bioinformatics Data Scientist
Jan. 2020 to Sept. 2020

- Created 3 **Python** scripts to retrieve data and create **Plotly interactive visualizations,** which cut the turnaround time for **troubleshooting** and **monitoring next-generation sequencing (NGS)** production-line processes in half
- Discovered the **root cause** for a time-sensitive issue on a new assay that resulted in batch failures impacting 500+ samples
- Write maintainable **SQL** and **Datalog** queries for 5 endpoints in internal libraries used for data warehousing
- Helped maintain **Flask** app for warehousing data in a Redshift **SQL** database used as the main data-source for **Tableau** dashboards
- Created 15+ **Tableau** dashboards, which enabled our operations team to carefully monitor new product launches

NEXTBEE MEDIA
San Mateo
Data Scientist
Sept. 2019 to Dec. 2019

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

- Designed the **SQL database model** and wrote the queries used to build customer profiles based on orders data
- **Segmented** thousands of customers into **tier groups** based on features identified through **domain knowledge** of **ecommerce**
- 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

METIS
San Francisco
Data Science Fellow
Apr. 2019 to June 2019

- **Completed 5 business-oriented data science projects** as part of an immersive 12-week program focusing on classical machine learning, database management, deep learning, and project design

BIOVERATIV, FORMERLY TRUE NORTH THERAPEUTICS
South San Francisco
Research Associate 2
Jan. 2017 to Mar. 2019

- Led **protein-engineering project** on of our lead antibody drug (BIVV009). Handled all **cloning and expression** of antibody variants and most of protein purification. **Collected binding and efficacy (KD and IC50) data** and performed **regression analysis**, revealing a log-linear relationship between KD and IC50 and making it easy 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. Used **PyMol** to visualize crystal-structure data returned from our CRO.
- Wrote **Python script** to automate design of short DNA oligos, which was over 200 times faster than manual design.
- Wet lab: PCR, tissue culture, protein expression and purification, ELISAs, affinity and kinetics characterization, crystallography

GENE YEO LAB, UCSD
La Jolla
Staff Research Associate 1
May 2013 to Nov. 2016

- **Developed standardized protocols** to transfect cultured HEK293 cells with viral expression constructs
- **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: cloning and expression of viral vectors, qPCR, CRISPR-Cas9, NGS-related pipelines

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

- Designed the **SQL database model** and wrote the queries to generate **Plotly interactive visualizations** all on the server side
- Enable users to **upload their own data** and generate their own figures
- Used **memory caching** to improve user experience and streamline data access

EDUCATION

University of California, San Diego — 2015
Double Major: B.S. Physics, B.S. Physiology & Neuroscience