HARRISON WANG

BIOINFORMATICS DATA SCIENTIST

harrison.c.wang@gmail.com
https://harrisonized.github.io/about
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
WEB FRAMEWORKS: Flask

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** production-line processes in half
- Created 15+ **Tableau** dashboards, which enabled our operations team to carefully monitor new product launches
- Write maintainable **SQL** and **Datalog** queries for 5 endpoints in internal libraries used for data warehousing
- Discovered the **root cause** for a time-sensitive issue on a new assay that resulted in batch failures impacting 500+ samples
- Performed **exploratory data analyses** to provide 10+ clients with answers for why their samples are failing
- Helped maintain **Flask** app for warehousing data in a Redshift **SQL** database used as the main data-source for **Tableau** dashboards

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/

- Write the **SQL** database model and 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 **time-series forecasting** to predict future revenue and number of new customers
- Used **logistic regression** and **random forest** to predict the likelihood of each customer making another purchase based on their purchase history
- Made 20+ **interactive dashboards** using **Plotly** and designed **UI mock-ups** using the InVision App

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

- **Completed 4 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

- **Completed 2 research projects** on the structural biology of our lead antibody drug. Independently designed and optimized experiments to test hypotheses.
- Performed **regression analysis** on protein-engineering data I collected in my experiments. Discovered a log-linear relationship between a physical property of our lead antibody drug and its efficacy at treating disease, making it easy to decide which drug variants to use in downstream experiments
- Wrote a **Python script** to **automate design** of short DNA oligos, which is over 200 times faster than manual design

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

- **Co-authored a Cell paper** that included my experiments on using a new genome-editing technology (CRISPR/Cas9) to track RNA in live cells
- **Co-authored a Neuron paper** that included my experiments on investigating the mechanistic link on how an RNA-binding protein could cause ALS
- Performed **production-line lab work** to prepare cell tissue samples for **next-generation sequencing**

FEATURED DATA PROJECTS

CLIMBING TRACKER WEB APP

Available at: https://harrisonized-climbing-app.herokuapp.com

- Built an **analytics dashboard** deployed to **Heroku Flask** to track personal climbing progress
- Executes **SQL** queries and generates **interactive visualizations** with **Plotly** all on the server side

ANALYZING YELP REVIEWS FOR CLIMBING GYMS

Available at: https://harrisonized.github.io/2019/06/05/yelp-climbing-gyms.html

- Used **multi-class classification** on user reviews to predict the number of stars given by the reviewer. Adjusted the class weights to give minority classes more importance, which improved my out-of-sample accuracy score from 0.635 to 0.867.
- Used **natural language processing (NLP)** techniques to **model topics** for 1-star and 5-star reviews.

EDUCATION

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