

HARRISON WANG

DATA SCIENTIST

✉ harrison.c.wang@gmail.com
🌐 <https://harrisonized.github.io/about>
in <https://www.linkedin.com/in/harrisonized/>
📱 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

Dry Lab Operations Data Scientist

San Francisco
Jan. 2019 to Current

- Perform **exploratory data analyses** to monitor and troubleshoot production-line **next generation sequencing processes**
- Perform **root cause data analyses** to rapidly diagnose time-sensitive issues arising from new product launches
- Write **reusable Python scripts** using Jupyter notebooks to **automate data retrieval**, **calculate statistics**, and **create interactive visualizations** using **Plotly** for presentation to stakeholders
- Write **maintainable Datomic datalog** and **SQL** queries for endpoints in internal libraries for **data warehousing**
- Use **Tableau** to build analytics dashboards for **monitoring metrics** for new products

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/>

- Write the **SQL** database model and queries used to build customer profiles based on orders data
- **Segmented customers** into tier groups based on features identified through **domain knowledge** of **ecommerce**
- Used the **time-series forecasting** method of SARIMA 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 **interactive analytics dashboards** using **Plotly** and designed **UI mock-ups** using the InVision App

METIS

Data Science Fellow

San Francisco
Apr. 2019 to June 2019

- Completed four **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

Research Associate 2

South San Francisco
Jan. 2017 to Mar. 2019

- **Completed 2 research projects** on the structural biology of our lead 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 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

Staff Research Associate 1

La Jolla
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** in **Plotly** all on the server side

SF RENT PETITION TIME SERIES FORECASTING

Available at: <https://harrisonized.github.io/2019/06/25/sf-rent-petitions.html>

- Used the **time-series analysis** method of SARIMA on counts of rent petitions in SF to forecast customer demand for up to 5 years
- Used unemployment rate as a feature in **linear regression** to improve training and prediction speed while achieving similar accuracy
- **Visualized** the distribution of rent petitions using **Tableau**

EDUCATION

University of California, San Diego

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

2015