# AN HOANG
## DATA SCIENTIST

## Experience

### Brava
Data Analyst

Redwood City
Aug. 2018 to Feb. 2019

- Designed and implemented experiments to synthesize data for predictive models and algorithms.
- Prototyped and tested models on oven hardware to build a smarter oven that will improve user experience and perfect the cooking process.

### UC Berkeley
Dance Coach

Berkeley
Oct. 2013 to Sept. 2018

- Designed a yearly dance syllabus and taught all levels of dancers from beginner to advance in order to train a nationally competitive dance team.
- Provided mentorship and guidance to create a self-sustaining team governance structure while sharing personal dance expertise and passion to enrich the experiences of student dancers.

### MapJam
Data Analyst

San Francisco
Apr. 2015 to Sept. 2015

- Provided product analytics by aggregating multiple sources of data into a SQL database and building out a frontend interface (hosted on AWS, using flask/jinja) for key statistic summaries.
- Conducted analytics as necessary in a startup environment, and attended ESRI startup convention with co-founder as product evangelist.

## Data Projects

### Pedestrian Dectection with Masked R-CNN
Proof of concept application of a state-of-the-art computer vision algorithm to count, segment, and colorize individual pedestrians in a scene using a masked R-CNN. Demonstrated how such an algorithm can be used for a wide variety of tasks such as counting foot traffic, intruder detection, or as a stylized 'color splash' photo generator.

### Satellite View of Inequality
Used GIS tools and a random forest algorithm to predict high poverty blockgroups in San Francisco.  Joined high resolution orthoimage data of SF from USGS's aerial photography asset with tabular data (corresponding to American Community Survey's income statistics) to train a classifier to predict the poverty level given a square aerial image patch within SF.

### Natural Language Processing on Stock Movement
Compiled financial filings from 12 years of 10-K and 10-Q data from companies within the Russell 3000 index along with numerical stock features to perform a time series prediction on the market -day-after-filing's stock pricing  (normalized to index performance). Models used pre-trained GloVe embeddings to transform textual data, which was ran on a CNN network, and a GRU network.

### Prius Price Prediction
Gathered data on a particular day for all of the United States from well known online used car service to predict the price of used Priuses. Used a set of ten primary features such as car age and odometer mileage in a linear classifier (with transformed features) to predict the price of used Priuses with a 83% holdout accuracy.

## Contact

- ✉ amhoang@gmail.com
- 🌐 newprior.com
- ☎ (562) 290-2866
- in a-m-hoang
- ⌂ redgeryon

## Education

### University of California, Berkeley
BA Japanese

### Deeplearning.ai
Deep Learning Specialization
Covered neural nets, CNN, RNN, LSTM, GRU, hyperparameter tuning, and practical ML project structure

## Skills

**PROGRAMMING:**
Python
Scikit-Learn
Statsmodels
Pandas
GeoPandas
R
SQL
mongoDB
Keras/Tensorflow
Rasterio/GDAL/Fiona
AWS/EC2
Spark

**DATA SCIENCE/ML:**
Bayesian Models
Regression
Classification
Sequence Models
Scraping
Neural Nets
Clustering
NLP