

# Zihan Ye

DATA SCIENTIST AT NATERA

☎ (+1) 718-751-6187 | ✉ zihanye96@gmail.com | 🏠 zihanye96.github.io | 📷 zihanye96 | 🌐 zihanye

## Experience

### Data Scientist

NATERA

San Carlos, CA

December 2020 - Present

- Utilized Python, R, and SQL to conduct exploratory data analysis, data visualization, and statistical modeling.
- Designed and executed experiment protocols for analytical validation studies to support product launch.
- Wrote production code in Python to parse data pipeline output and efficiently compute and visualize quality control metrics.
- Communicated data insights and recommendations to cross-functional stakeholders to influence business decisions.

### Product Marketing Analyst

MBI, INC.

Norwalk, CT

August 2018 - August 2019

- Queried and analyzed data on market segments and customers to develop product strategy and achieve industry-leading profit margins.
- Managed competing project timelines of 70+ consumer products totaling over \$1MM in sales with advertising expenditures of over \$200M.
- Led cross-functional teams to conceptualize and execute creative initiatives such as product development and marketing materials.

## Education

### University of California, Berkeley

MASTER'S IN STATISTICS

Berkeley, CA

August 2019 - May 2020

- Relevant Coursework: Applied Machine Learning, Advanced Probability, Mathematical Statistics, Linear Models, Statistical Computing
- Cumulative GPA: 3.90/4.00
- Awards: Outstanding Graduate Student Instructor Award, Betty Scott Excellence Scholarship

### Williams College

BACHELOR'S IN STATISTICS

Williamstown, MA

August 2014 - June 2018

- Relevant Coursework: Discrete Mathematics, Linear Algebra, Real Analysis, Abstract Algebra, Data Mining, Bayesian Statistics, Statistical Inference, Regression Theory, Categorical Data Analysis, Game Theory
- Cumulative GPA: 3.79/4.00
- Awards: Mu Sigma Rho (Statistics Honor Society), Dean's List

## Projects

### Spotify Data Visualization (Python)

- Utilized Spotify API to download personal account data and summarized music preferences and behavioral patterns via data visualization.
- Created classification models to recommend music based on my Spotify streaming history and playlist data using XGBoost and TensorFlow.

### Toxic Comment Classifier (R)

- Constructed various classification models to correctly identify toxic online comments over 93% of the time (Naive Bayes, Random Forest, Logistic Regression, Neural Network, SVM).
- Utilized feature engineering techniques and problem context to develop 11 intuitive features used for modeling.

### Airbnb Income Predictor (Python)

- Developed machine learning models to accurately predict annual income for listings on Airbnb.com using web-scraped data.
- Researched existing literature and utilized data cleaning and feature engineering techniques to prepare dataset for modeling (OLS, LASSO, Random Forest, GBM) and inference.

## Skills

**General** Machine Learning, Data Visualization, Statistical Modeling, A/B Testing, Experimental Design, Data Mining

**Technical** Python, R, SQL, Excel, Bash, git, Tableau, Spark, Hive, AWS, Jupyter

**Libraries** Numpy, Pandas, SciPy, Matplotlib, scikit-learn, PySpark, XGBoost, TensorFlow, Keras, dplyr, ggplot2, Shiny