

# Selina Wang

857-334-2023 | selina\_xinyue\_wang@outlook.com | [Github](#)

## EDUCATION

---

### Master of Science in Data Science

**Brown University**

GPA: 3.80

**Relevant Courses:** Machine Learning (Python), Deep Learning (Python), Data Engineering (SQL), Statistical Learning (R), Data Visualization (R, Adobe Illustrator)

May 2024

Providence, RI

### Bachelor of Science in Mathematics and Quantitative Economics

**Tufts University**

GPA: 3.95

**Relevant Courses:** Mathematical Modeling, Econometric Analysis, Calculus, Statistics

May 2021

Medford, MA

## PROFESSIONAL EXPERIENCE

---

### Data Science Intern

**Citizens Bank**

May 2023 - December 2023

- Designed and developed machine learning and data pipeline in Python with a team of 5 using Git to automate model training on AWS cloud, reducing 10 hours of labor work per week
- Cleaned, preprocessed and performed exploratory data analysis on transaction datasets with 9M+ observations in Python to develop a predictive machine learning model for fraud detection
- Integrated and transformed data from various sources to ensure data quality and consistency; communicated with data engineers to investigate data discrepancies
- Collaborated cross-functionally with strategy teams and business stakeholders; communicated key insights to a non-technical audience through presentations and data visualizations

### Data and Analytics Intern

**Shanghai Roots and Shoots — Million Tree Project**

October 2021 - February 2022

- Performed statistical analysis in Python for 20 different plantations to uncover trends in biodiversity data and identify plantations with irregular growth patterns; visualized results using Tableau
- Implemented classification algorithm using Python to analyze image data and estimate plant coverage rates, increasing accuracy by 30% compared to baseline approach

### Intern

**Industrial Economics, Inc. (Environmental Economic Consulting)**

May 2020 - August 2020

- Analyzed the impact of rising Sea Surface Temperatures on vibriosis (an ocean bacterial infection) using a linear regression model in STATA
- Co-authored the research paper *The influence of climate change on vibriosis in the United States: projected health and economic impacts for the 21st century* published on *Environmental Health Perspectives*

## DATA PROJECTS

---

### Deforestation and Commodity Consumption Data Narrative

March 2024 - May 2024

- Crafted data visualizations to uncover the relationship between US beef consumption and deforestation abroad using ggplot and Adobe Illustrator; hosted data narrative on website using HTML

### Pharmacy Database Design and Implementation

October 2023 - November 2023

- Designed a relational database in SQL; implemented constraints and triggers to ensure consistency and executed SQL queries to extract information about drug prescriptions

### Ground Water Time-Series Modeling Challenge

October 2022 - December 2022

*A challenge hosted by the European Geoscience Union*

- Collaborated with team member to build a regression model that predicts water levels in wells using meteorological data, achieving an MSE of 0.027

## SKILLS

---

**Programming Languages:** Python (Scikit-Learn, TensorFlow, Keras, Pandas, NumPy, Matplotlib), R, SQL, Hadoop, Spark, Seaborn, Plotly, ggplot2

**Data Science & Miscellaneous:** Data science pipeline (cleaning, visualization, feature engineering, modeling, interpretation), Excel VBA, Git, AWS Sagemaker and S3, Tableau, Power BI, A/B Testing

## CERTIFICATIONS

---

**Google Data Analytics Professional Certificate**