

# MATTHEW SAO

+1(415) 933-7871 ◊ San Francisco, CA

[msao19981@gmail.com](mailto:msao19981@gmail.com) ◊ [linkedin.com/in/matthew-sao](https://www.linkedin.com/in/matthew-sao) ◊ [www.github.com](https://www.github.com) 

## EXPERIENCE



---

### Data Intern

Sept 2020 - Present

U.S. Department of the Interior - Office of Natural Resources Revenue

10 hours/week Part-Time Remote

- Verified and prepared production, revenue and disbursement data for [revenue.data.doi.gov](https://revenue.data.doi.gov) 
- Helped to design and implement new visualizations such as our renewables fact sheet, taking into account accessibility concerns and visualization best practices.
- Updated data documentation as well as internal data checklists to improve the data review process. 

### Quantitative Analytics Intern

Jan 2022 - April 2022

U.S. Securities and Exchange Commission - Quantitative Analytics Unit

15-20 hours/week Part-Time Remote

- Constructed a corpus of text documents from UNPRI public signatory reports using Python and PDFMiner.
- Created functions to extract survey information from semi-structured text documents and performed preliminary clustering based on term frequency.
- Analyzed market opening and closing orders using time-series analysis.

### Analytics Intern

May 2020 - Aug 2020

U.S. Department of Veteran Affairs - Digital Media Engagement

10 hours/week Part-Time Remote

- Utilized Salesforce Social Studio to analyze metrics for the Department's five social media channels.
- Used data to provide insight on audience reception of content with engagement metrics such as likes, shares, comments, etc.
- Utilized Google Analytics to summarize and compare metrics such as page views on a monthly basis for the Department's official blog, VAntage Point.

## PROJECTS

---

### Political Popularity of Misinformation

[GitHub](#)

- Senior capstone project that aimed to analyze the growth rates of politicians on Twitter who spread misinformation and ones who did not.
- Used Python to create a data pipeline to obtain tweets and engagement metrics from Twitter's API.
- Created popularity estimates to analyze reception to tweets and used statistical methods to analyze the growth of politicians over time.

## EDUCATION

---

**Master of Computer Science**, University of Illinois Urbana-Champaign

Anticipated May 2023

GPA: 3.86

Relevant Coursework: Natural Language Processing, Applied Machine Learning, Data Mining, Data Visualization, Applied Statistics, Distributed Systems, Practical Statistical Learning

**B.S. Data Science**, University of California, San Diego

March 2021

Relevant Coursework: Systems for Scalable Analytics, Statistical Methods, Machine Learning, Probabilistic Models, Data Structures and Algorithms, Data Visualization, Data Analysis and Inference

Provost Honors: Winter 2019, Winter 2020, Spring 2020, Fall 2020

## SKILLS

---

**Languages** Python, R, Java, SQL, Javascript, HTML, CSS

**Tools** Git, Docker, AWS 