

CONTACTS

sara.zong@gmail.com
(415)823-7056
San Francisco, CA 94122
[LinkedIn: sarazong](#)
[Github: sarazong](#)

SKILLS

LANGUAGES

Python
R
SQL

MACHINE LEARNING

Numpy
Pandas
Scikit-learn
Selenium
BeautifulSoup
NLTK

VISUALIZATION

Matplotlib
Seaborn
ggplot
Tableau

DATA BASES

PostgreSQL

EDUCATION

University of Michigan, Ann Arbor
M.S. Biostatistics, 2020

University of California, Davis
B.S. Molecular Biology, 2008

EXPERIENCE

Data Scientist

Jan 2021 – Mar 2021

Metis

Completed an immersive 12-week data science bootcamp with a strong emphasis on project-oriented skill-building in project design, data wrangling, visualization, statistical modeling, machine learning, and communication of findings. Project highlights include:

Content-based Non-fiction recommender:

- Scraped data for 3000+ non-fiction books using Selenium and BeautifulSoup
- Performed natural language processing and topic modeling on book summary
- Built a content-based book recommender and deployed on Streamlit.io

Podcast Recommendation Based on Reviews:

- Performed text analysis with natural language toolkit (NLTK) and topic modeling with Scikit-learn on podcast reviews to identify topics
- Built a podcast recommendation system based on listener reviews

Predicting Flight Delay:

- Applied both logistic regression and Random Forests to flight performance data obtained from the Bureau of Transportation Statistics to predict flight delay status

Predicting San Francisco Housing Prices:

- Scraped 1,000+ San Francisco house listings using Selenium and BeautifulSoup
- Modeled house data with vanilla linear regression and Lasso linear regression to predict housing prices

Research Assistant

Sep 2019 – Jun 2020

University of Michigan, Ann Arbor

- Examined the association between single nucleotide polymorphism of certain genes and the side effect of breast cancer treatments with logistic regression
- Generated figures for publication

Graduate Student

Aug 2018 – May 2020

University of Michigan, Ann Arbor

Completed a capstone course focused on project-oriented data analysis and statistical inference, with clinical and social study data from past or on-going studies.

Sample projects include:

Study transition from cigarette to e-cigarette:

- Examined the relationship between flavored cigarette and smoker's transition to e-cigarette by using logistic regression with sample design incorporated into the analysis

Survival analysis of head and neck cancer:

- Studied the association between head and neck cancer outcomes and common medications use for chronic conditions
- Wrote up statistical analysis reports

Senior Research Associate II

Jan 2015 – Aug 2018

Assembly Biosciences, South San Francisco, CA

- Developed biological assays to capture the different populations of hepatitis B virus in clinical samples
- Performed experiments, analyzed data, and contributed to publications
- wrote standard operation procedure for assays developed and trained junior research associates