

Ming Senn Teo

628-758-5826 | mingsennteo@berkeley.edu | [linkedin.com/in/mstteo](https://www.linkedin.com/in/mstteo) | github.com/MSTeo01 | <https://msteo.org/>

EDUCATION

University of California, Berkeley

Berkeley, CA

Bachelor of Arts in Data Science (Emphasis in Economics)

May 2024

Relevant Coursework: *Principles and Techniques of Data Science, Data Mining and Analytics, Data Engineering, Data Structures, Probability for Data Science, Data Science for Economists, Data Inference and Decisions, Econometrics*

De Anza College

Cupertino, CA

Associate in Arts - Economics for Transfer, Associate in Science - Mathematics for Transfer

June 2022

TECHNICAL SKILLS

Languages: Python, SQL, R, Java, C++, Javascript

Databases: PostgreSQL, MongoDB, MySQL

Code Management: Git/Github, Jupyter Notebook, Visual Studio, IntelliJ, RStudio, PyCharm

Libraries: Pandas, Numpy, Scipy, NLTK, Pytorch, Scikit-learn, Tensorflow, Keras, Matplotlib, Seaborn

EXPERIENCE

UC Berkeley School of Public Health

Berkeley, CA

Data Analyst

September 2023 – May 2024

- Improved runtime efficiency by 30% through optimizing PostgreSQL queries on massive (order of trillion rows) production databases
- Built causal inference models for prices charged by recently acquired healthcare giants to prompt California's healthcare policy changes
- Visualized prices data using matplotlib and seaborn to identify trends and outliers between certain insurers on a subset of medical procedures

De Anza Palgorithm

Cupertino, CA

Peer Tutor

September 2021 – June 2022

- Planned and facilitated workshops aimed at introducing fundamental programming concepts to new members
- Provided 1-1 guidance to members on questions relating to debugging and conceptual help working on Python projects, which includes programming fundamentals and building simple regression models
- Collaborated with officers in developing written materials for the workshops that align with the members' learning needs and goals

PROJECTS

Capstone Project - Causality of Arkansas Speed Limit Increase on Traffic Accident and Effect of Road Conditions on Traffic Accident Duration |

Spring 2024

Python, Numpy, Pandas, Scikit-learn, Matplotlib, Statsmodel, Seaborn

- Built causal inference models at both state and county level that accounted for weather and location, showing the significance of the limit raised in traffic accidents
- Trained a random forest model that combined the disparate sources into one projection that outperformed the mean squared error of a Frequentist GLM model by 12%
- Concluded that Arkansas Speed Limit Increase has a significant impact on the number of traffic accidents and that certain road conditions serve as important factors for predicting accident duration

Spam Email Classifier | *Python, Pandas, Scikit-learn, Numpy, Matplotlib, Seaborn*

Spring 2023

- Trained a logistic regression model using scikit-learn to classify spam emails based on keywords and email length with a validation accuracy of 94%
- Implemented NLP practices into feature engineering to improve the baseline model accuracy by 20%

LEADERSHIP

De Anza Palgorithm

Cupertino, CA

Secretary and Inter Club Council Representative

September 2021 – June 2022

- Co-managed recruitment for Spring 2022 quarter which included leading outreach events and conducting behavioral interviews for officer positions
- Recorded, maintained, and managed accurate minutes of all club meetings
- Prepared and distributed biweekly agenda notes on future campuswide club policies and activities to club members