# Wai Yam (Gordon) Tsai

Berkeley, CA| 925-658-8921 | tsaiwaiyamgordon1126@gmail.com

## EDUCATION

**Stanford University** **Stanford, CA**
*M.S. Management Science of Engineering* Start at Sep 2022
**University of California Berkeley** **Berkeley, CA**
*B.A. Data Science – GPA: 3.8* Grad. in May 2022
Relevant Coursework: Probability, Statistics., Computer Science, Data Science & Modeling, Machine Learning, Data Optimization, Business Analytics, Linear Algebra, Abstract Algebra

## WORK EXPERIENCE

**Data Analytics Research Lab | University of California Berkeley** **Berkeley, CA**
*Research Assistant* July 2021 – Present

- Cleaned, transformed, and merged 100,000+ rows of data and implemented regression model and feature engineering from scratch for the fleet management optimization of electric trucks
- Identified the shortest path algorithm, provided the animation of the vehicle movement to predict the warning level of the current situation, for example using logistic regression to predict the probability of brake failure
- Used principal component analysis (PCA) to reduce the feature space, and leveraged multiple linear regression to forecast the battery level for electrical trucks
- Optimized Fleet Management time using the battery level forecasts and warning level prediction to reduce the cost of the Fleet Management team to a minimum level

*Textbook Editors (Data Science for Business Decision)*

- Started with the item-based collaborative filtering for movie recommendation, which is based on KNN and Cosine similarity for nearest neighbor search to avoid "Curse of Dimensionality".
- Improved the movie recommender by utilizing Alternating Least Square (ALS) Matrix Factorization to overcome the shortcomings of popularity bias and item cold-start problem.
- Conducted model hyper-parameters tuning with ML cross-evaluation toolbox and monitored data processing performance via Python.

**Mangul Research Lab | University of Southern California Clinical Department** **Los Angeles, CA**
*Research Assistant for HLA Data Analysis* July 2021 – Present

- Utilized ANOVA to test the statistical significantly different between Healthy and Disease groups thru RNA pattern
- Implemented Random Forest model and Cross-Validation to predict certain individual variations in a person's immune system based on clinical syndromes and RNA sequences achieved 85%+ accuracy
- Conducted in-depth research on tests results and summarized findings for the research of Prediction of Next Generation Sequencing Research on HLA and RNA Sequences

**Hong Kong Trade Development Council** **Hong Kong**
*Data Analyst intern, Database intelligence* Jun 2019 – Aug. 2019

- Completed administrative activities such as supporting business departments for marketing, expanding the TDC database, and systemizing database records
- Utilized mathematical knowledge in creating pragmatic solutions to determine the true value of data
- Enhanced the business system by data analysis, resulting in more expos and events partnership opportunities

## ACADEMIC PROJECT

**Causal inference in Bidens Elections** **Berkeley, CA**
*Final-year project* Mar 2022– May 2022

- Applied causal observational study to discover the causal relationship between the number of unique donations a president candidate received and the proportion of the vote they received in the primary.
- Constructed Causal Directed Acyclic Graphs (DAG) to reveal the causal relationship between the treatment and the outcome, and used the Pearl's Back-Door criterion to determine the minimum set of confounding factors.
- Performed OLS regression adjustment, inverse probability weighting and propensity score matching to tear out the effect of confounding factors, and estimated the causal impact of number of donations on proportion of vote received in the primary.

## SKILLS & INTERESTS

**Programming/Software:** Python, R, SQL, Java, Tableau, Advanced Excel, Matlab
**Statistics Analysis:** A/B Testing, Hypothesis Testing, Time Series Analysis, Feature Engineering
**Machine Learning:** Regression, Decision Tree, Random Forest, Clustering, KNN, SVM, Naive Bayes Algorithm