# DANIEL HUANG

daniel.n.huang@gmail.com ◇ (626) 215-7238 ◇ github.com/dnhuang

## EDUCATION

**University of California, Berkeley** *Fall 2022*
Bachelor of Arts, Data Science (Emphasis: Economics)

## EXPERIENCE

**The Entrust Group**
*Real Estate Associate* *July 2023 - January 2024*

- Reduced hundreds of manual work hours to a single program run by designing and implementing an ETL pipeline to automate the generation of real estate transaction reports adherent to strict business requirements.
- Provided insight on company's key performance indicators to directors and managers at annual all-hands meeting with curated visualizations and dashboards, enabling data-informed decisions.
- Further improved company efficiency by writing customized scripts and programs that automated various administrative tasks and ensured they were accessible to users without extensive technical backgrounds.
- Leveraged knowledge in Python (tkinter, pypdf, pytesseract), OCR, SQL, VBA, zsh, Microsoft suite

**UC Berkeley Division of CDSS & UC Berkeley EECS Department**
*Computer & Data Science Courses TA (CS 61A, 61B, C88C, Data 100, Physics 88)* *Aug 2021 - Dec 2022*

- Instructed diverse groups of over 50 students on data structures and algorithms, and collaborated with course staff to supervise office hours, labs, and discussions to improve student outcomes.
- Developed and iterated on educational resources across various data and computer science courses by refining code scaffolding to better align with learning objectives.
- Leveraged knowledge in Python (pandas, matplotlib, seaborn, scikit-learn), Java, data structures, algorithms, git

## PROJECTS

For additional projects & repositories: github.com/dnhuang

**Fifa World Cup 2022 Predictor**

- Leading a team of seven, took the initiative to delegate and monitor tasks, maintaining clean version control and minimizing merge conflicts by enforcing the use of pure functions.
- Scraped world cup records from a wide range of sources; cleaned, consolidated, and transformed the data to train an ensemble model; predicted **Argentina**, Brazil, and Italy as the potential winners.
- Utilized: Python (pandas, seaborn), logistic regression, random forests, neural nets, boosting

**Relationship Between Mobility and US GDP and Investigating Causes of Traffic Fatalities**

- Found changes in GDP were correlated with increases in residential mobility using multiple hypothesis testing and confirmed significance of discoveries with naive, Bonferroni, and B-H correction.
- Noted a relationship between traffic fatalities and infrastructure spending by constructing a Bayesian hierarchical model and contrasting it with GLMs.
- Utilized: Python (pandas, scikit-learn, PyMC3), data science significance testing and modeling techniques

**Gender Proper Case Usage**

- Inspired by a debate among friends, investigated the difference in "proper case usage" in messages between genders and established statistical significance through A/B testing; peripherally uncovering a gender hiring gap.
- Scraped publicly available "bio" data from CS 61A course staff from a range of semesters to perform text analysis.
- Utilized: Python (pandas, matplotlib, seaborn, scipy, regex), A/B testing, statistics

**Spending Analysis**

- Identified a personal tendency towards food expenditures (46% of whole) by consolidating and analyzing credit card transaction data from all bank accounts and summarizing the information in an interactive dashboard.
- Utilized: Python (pandas), SQL, Tableau

## SKILLS

**Programming:** (proficient): Python, Java, SQL, git (familiar): C, Go, JavaScript, HTML/CSS, zsh, VBA, Matlab
**Data Science:** hypothesis testing, Bayesian hierarchical modeling, generalized linear models, machine learning