Zelong/Alan Wang

Contact

Address

La Jolla, CA, 92122

Phone

858-539-5564

E-mail

zew013@ucsd.edu

GitHub/Personal Web

zew013.github.io

Test Taken

GRE Verbal: 162

GRE Quantitative: 169

Skills

Python — Proficient

Stata — Very Good

R — Good

Microsoft Office Suit — Good

Java — Average

MATLAB — Average

Git — Average

Tableau — Average

HTML — Some Experience

Mandarin — Excellent

Time management — Excellent

Will learn SQL/NoSQL and tools such as MapReduce/Hadoop and Spark systematically this Fall/Winter

Education

The University of California, San Diego | 2020-Present | GPA:3.98

B.S | Data Science

Related Coursework: Python, Java, HTML, Data Structures, Data Science Algorithms, Web Scraping, Data Processing, Visualization, Machine Learning, NLP basics, Deep Learning basics.

B.S | Joint Major in Math and Economics

Related Coursework: Calculus, Linear Algebra, Numerical Analysis, Optimization, Probability and Statistics, Microeconomics, Econometrics (T, Z, Chi-squared, F test, Regressions, experiment design, panel data, etc.), Operations Research.

Experience

Research Assistant | Professor Dale Squires and Professor Richard Carson | Aug. 2022 - Present

- Explore the underlying pattern in terms of international organizations and their membership fees.
- Extract and prepare data from low quality UN reports (1950-2022) using Python, SOAP OCR API, and AWS.

CSE-PACE Program Designer | UCSD | Apr. 2022 – Sep. 2022

- Searched and reviewed a miscellaneous assortment of papers, articles, programs, and games. (approx. 30 pages/week)
- Design CSE cohort programs and build 7 courses on various CS/DS related topics.

Research Assistant | Professor Richard Carson | Dec. 2021-Present

- Did literature review and gained background knowledge on discrete choice models and related topics.
- Arranged and corrected research data to create representative graphs intended for academic paper. (approx. 60 graphs)
- Coded Monte Carlo Simulation in Stata. Performed qualitative analysis. Coded to achieve conditional logit model.

Data Analyst/Tech VP | Lumnus Consulting (Student Enterprise) | Nov. 2021-Present

- Collaborated with business-unit leaders to identify and prioritize problems.
- Built prediction models based on historical Instagram and Twitter data. Created visualizations and analyses for the team. Delivered project presentations on how we should present our organization on social media.
- Organized team events and fostered team communication.

Data Manager & Builder | Robotics Team | 2018 – 2020
Tutor & Math Team Leader | 2019 - 2021

Self-Driven Projects

Forage Data Analytics Virtual Experience programs | Python, Tableau, Excel, Git

- Accenture: After cleaning data anomalies and merging data in **Python**, created **Tableau** dashboard that demonstrates relationships between key features. Practiced storytelling and **presentation** skills by delivering our findings.
- BCG: Deployed automatic EDA using **SweetViz**. Defined, assigned price sensitivity to each customer. Tested hypothesis by **permutation test**, **K-S test**, and heat map. Predicted customer churn by **Random Forest** and **XGBoost**.

Analysis of Power Outage Status in the Continental U.S. | Python, Excel

- Went through the full process of questioning, researching, data cleaning & EDA, missingness assessment, hypothesis test, baseline model, more feature engineering, final model, and fairness analysis.

Indian Crop Production and Indian Climate Analysis | Python, Slides

- Literature review on previous reports about Indian agriculture, pastoralism, and climates.
- Data gathering, processing, interactive visualization using **bs4**, **JSON**, **Geopandas**, **Altair**, **Folium**, **Plotly**.
- Reported findings and provided suggestions.

Simple Language Model | Python

- **Web scraping** for specific books from a public book website. Tokenize corpora.
- Created an N-Gram Language Model that can generate paragraphs resembling the style of an author or a book.

Wealth Prediction based on 1991 Survey of Income and Program Participation (SIPP) | R

- Gained background knowledge about data. Applied EDA, feature engineering.
- Incorporated and compared **polynomials** and **splines** in **GAM**. Tested **full OLS**, **Lasso**, **Ridge**, **Stepwise regression model** and **Random Forest model**.