GUOXUAN(Jason) XU

g7xu2024@gmail.com | (909) 760-9522 | La Jolla, San Diego, CA

EDUCATION

University of California San Diego, Bachelor of Science, Data Science **Jun 2022 – Jun 2026 (expected)**

- GPA 4.0/4.0
- Provost Honors
- Course Work: Data Structure, Data Management, Cloud Computing, Probability Theory, Modeling, Linear Algebra, Recommendation System

Walnut High School, Walnut, CA, High School Diploma **Jun 2018 – Jun 2022**

- GPA 3.93/4.0
- English Excellence Award (sole recipient)

SKILLS

- Coding: Python, Java, SQL, R, Git
- Statistics Inference: Hypothesis Testing, A/B Testing, Probability, Linear Regression, Logistic Regression, Time Series, Fairness Analysis
- Machine Learning: Data Cleaning, Feature Engineering, Scikit-learn, Random Forest, Pandas, Numpy, Grid Search, Web Scrapping
- Data Visualization: Tableau, Matplotlib

EXPERIENCE

Data Analytical Intern, Tong Consulting Inc. **Aug 2024 – Present**

- Interacted with client to recommend 5 profitable round-trip routes and KPIs
- Processed encrypted flights data using Python for data cleaning and munging
- Configured linear regression predicting future flight delay in minute with RMSE of 8.98
- Delivered analytical mythologies, insights, and recommendations to client through PowerPoint and reports

Student Research Intern, The Wu Lab at Scripps Research **Jun 2024 – Aug 2024**

- Performed data imputation to improve the quality of raw biomedical data
- Used Pandas and NumPy for data cleaning, hypothesis testing, modeling and downstream application
- Configured multi-class logistic regression with F1-score of 0.74 on data imputation
- Presented findings via poster and PowerPoint to **50+** interdisciplinary biomedical researchers

Instructional Assistant, UCSD Halıcıoğlu Data Science Institute **Sep 2023 – Jun 2024**

- Maintained course grading pipeline that calculated and displayed students' grading summaries
- Taught **500+** students data science principle and application mainly using Pandas and NumPy
- Developed python scripts that calculates students' grades and display grading summaries
- Deployed transparent grading pipeline for **20+** students per course

PROJECT

Predicting Electricity Consumption, UC San Diego Course Project **Oct 2023 – Nov 2023**

- Performed open-ended data analytical investigation on historical US power outage events
- Used Pandas and NumPy for data cleaning, exploratory data analysis, and statical testing
- Created Random Forest Regressor forecasting electricity consumption of US states
- Configured Final model with 0.14 mean absolute percentage error on testing data