Lou Zhou

Houston, TX
Email: lz80@rice.edu **Linkedin:** www.linkedin.com/in/lou-zhou/ **Website:** lou-zhou.github.io/ U.S. Citizen

Education

Rice University **Expected Graduation May 2027**

Bachelors of Arts in Statistics and Sport Analytics(3.89 GPA) *Houston, TX*

Relevant Coursework: Advanced Sport Analytics, Introduction to Program Design, Linear Regression, Real Analysis, Advanced Statistical Methods, Stochastic Methods, Tools and Models for Data Science, Practical Machine Learning

Languages & Tools: Python(scikit-learn, pandas, NumPy, PyTorch, Keras, BeautifulSoup4), R(tidyverse, ggplot, face, glmnet, shiny, caret), XGBoost, Java, Regex, SQLite, Git, EC2, EMR, Hadoop, PySpark

Experience

Rice Athletics **August 2025 - Present**

Student Technical Projects Co-Lead | Git, Python, Pandas *Houston, TX*

- Co-leading a 13-person engineering team delivering internal tools for Rice Athletics such as building a Python/Pandas automation tool that generates post-match reports
- Liaison to non-technical stakeholders, scoping features with scouting leads and translating requirements into clear specs and timelines in plain language

Rice University **August 2024 - Present**

Research and Teaching Assistant | Python, pandas, NumPy, PyTorch, Git, Matplotlib, SQLite *Houston, TX*

- Conducting research under Dr. Scott Powers in partnership with TSG Hoffenheim's Research Lab, evaluating how virtual reality training impacts in-game decision-making in soccer
- Reduced data processing time by 83% using pandas by redesigning a data processing pipeline that generated features from 300+ matches of soccer tracking and event data
- Expanded data availability for ML training by broadening data pipeline support to 3 new soccer data provider formats(SkillCorner, Sportec, Hawkeye) with pandas/NumPy
- Refined soccer pass quality estimations, adding features(e.g. player velocity) to a PyTorch CNN to allow the model to more accurately capture game context
- Achieved a 94% student satisfaction rate from semester survey by guiding 17 students as the sole TA in an applied data science course through in-class help and office hours

Carnegie Mellon University **June - July 2025**

Undergraduate Researcher(CMSACamp) | R, tidyverse, ggplot, Git, XGBoost *Pittsburgh, PA*

- Fully-funded research experience in sports analytics under the mentorship of Dr. Karim Kassam and Quang Nguyen
- Matched public research deep learning accuracy (59.9% vs. 59.8%) on NFL throw target prediction using simpler XGBoost model with two novel features describing potential separation and QB vision

St. Jude Children's Research Hospital **June - July 2024**

Biostatistics Research Assistant | R, tidyverse, ggplot, face, plink *Memphis, TN*

- Created ALS patient-specific disease evolution projections by building models to forecast ALS progression using functional PCA on longitudinal patient data in R
- Contributed to understanding of ALS genetic risk factors by identifying 3 progression markers with Genome-Wide Association Studies on genotype and PCA data using HPC resources

Projects

gg-pyscraper: A Python Library for Parsing Esports Data **July 2025 - Present**

- Expanding public ease of access to esports data by building a generalized BeautifulSoup4/regex parsing library that standardizes community-wiki pages into structured datasets for 55+ video games
- Work to be presented at the 2025 Carnegie Mellon Sports Analytics Conference

Team Coin Flip: Travel Fatigue and Performance **January 2024**

- Created XGBoost models with scikit-learn and a modified ELO ranking to assess travel impact on performance, placing 2nd of 59 teams at the 36-hour 2024 Rice Datathon

Breaking the Cycle: Reducing Recidivism in Iowa State Prisons **August 2022 - May 2023**

- Improved recidivism risk prediction by building a Keras neural network (0.85 AUC-ROC) to predict inmate recidivism probability, outperforming baseline models (0.64 AUC-ROC)
- Delivered economic burden estimates and policy guidance using Monte Carlo simulations and SHAP values in Python to model variability and recidivism risk factors
- Awarded 2nd place of 227 teams in the 2023 Modeling the Future Challenge, earning a $15,000 team prize

Additional

Activities: · *Tracks Organizer*, 2025 Rice Datathon · *Mentee*, 2025 MIT Sloan Sports Analytics Mentorship Program