STANLEY OU

301-273-8205 | stanleyo@andrew.cmu.edu | Burtonsville, MD | Portfolio | LinkedIn | GitHub

EDUCATION

Carnegie Mellon University (CMU)　　　　　　　　　　　　　　　　　　　**Pittsburgh, PA**
B.S. in Statistics and Machine Learning, Additional Major in Computer Science　　　*Aug 2022 – May 2026*
GPA: 3.83/4.0, **Honors:** Dean's List (all semesters)
- *Relevant coursework: Machine Learning, Data Science in R, Statistical Visualizations, Probability and Statistical Inference, Software Engineering, Artificial Intelligence, Computer Systems, Algorithms & Data Structures in C*

Montgomery Blair High School - Science, Mathematics, Computer Science Magnet Program　　**Silver Spring, MD**
SAT Score: 1560, GPA: 4.0/4.0 (4.82 weighted)　　　　　　　　　　　　　　　　　*Aug 2018 – Jun 2022*

SKILLS

- Programming Languages: Python, SQL (PostgreSQL, MySQL), R, C, Java, HTML, CSS, JavaScript
- Technologies: GitHub, Google Cloud Platform, Flask, Docker, Conda, Shiny, Tableau, Pandas, NumPy, Sklearn, Tensorflow, Pytorch
- Other: LLMs, Data visualization, Predictive analytics, OOP, Web development, Unit testing, Microsoft Office (Excel, Word, etc)

PROFESSIONAL EXPERIENCE

Ploomber, *Software Engineering Intern*　　　　　　　　　　　　　　　　　　**Remote**
　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　Jul 2024 - Present
- Use Python, Javascript, and CI/CD to drive the full-stack development of Ploomber's cloud data products.
- Write technical documentation and blog posts on implemented features and cloud deployment examples.
- Collaborate with a team of engineers to design and implement scalable cloud infrastructure, utilizing frameworks such as React, and deploying on AWS.

PA 211, *Data Science Intern*　　　　　　　　　　　　　　　　　　　　　　**Lemoyne, PA**
　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　Feb 2024 – Present
- Supported PA 211's statewide initiative in assisting thousands of people in need of shelter, food, and utilities by developing BI dashboards and data reports in DOMO to help guide 211's resource allocations.
- Created ETL pipelines using SQL to integrate data from external software and web-scraped data using scrapy.
- Developed a multinomial logistic regression model with 75% accuracy in predicting clients' identified needs.

CMU, *Undergraduate Teaching Assistant*　　　　　　　　　　　　　　　　　**Pittsburgh, PA**
　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　Aug 2023 – May 2024
- Aid the professor in grading and teaching 36-219 – Probability and Random Processes.
- Graded and hosted problem set sessions for 80-101 – Dangerous Ideas in Science and Society.

CMU LearnLab, *Undergraduate Research Assistant*　　　　　　　　　　　　　**Pittsburgh, PA**
　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　Jun – Aug 2023
- Meta-analyzed data collected from seven semesters. Calculated effect sizes, confidence intervals, z-scores.
- Cleaned, validated, and transformed data (log-odds conversion) in Excel to prepare for analysis.
- Created scripts for data visualizations in Jupyter Notebooks using Pandas, Matplotlib, and Seaborn.

UMBC Bennett Labs, *Research Intern*　　　　　　　　　　　　　　　　　　**Baltimore, MD**
　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　Jun 2021 – Jan 2022
- Performed data analysis on sulfur-based adsorbates' interactions with calcite and alumina surfaces.
- Used Linux-based systems for file retrieval, DFT calculations, and xmGrace to plot surface adsorption data.
- First-authored a paper published in *Clays and Clay Minerals*.

RELEVANT PROJECTS

Pandas Strftime Open-Source Contribution
- Implemented and submitted a pull request that resolves a discrepancy in pandas' DatetimeIndex.strftime method, which had inconsistent behaviour when called on single Datetime objects vs Datetime object arrays.
- Introduced an 'errors' parameter with values "raise", "warn", and "ignore" to the strftime method that offers users control over error handling and provides clearer error messages.
- Wrote unit tests that pass GitHub actions and contributed to pandas library's documentation.

Stock Portfolio Tracker and Predictor　　　　　　　　　　　　　　　　　　　Project Link
- Created a web application using OOP and Flask that enables users to track their portfolio performance and project future portfolio growth.
- Used yfinance web scraping API to retrieve stock price data and Pandas for data pipelining and feature engineering.
- Implemented a recurrent neural network using LSTM layers to perform time-series forecasting with a accuracy of ~70% on stock prices.
- Created a styled interactive webpage in HTML and CSS with embedded statistical plots created in Bokeh and portfolio statistics.

NBA Career Outcomes Predictor
- Built a Random Forest Classification model to predict NBA players' career outcomes based on their performance in their first four seasons, achieving a testing set accuracy of approximately 70%.
- Performed data cleaning and engineering techniques, including exploratory data analysis, merging datasets, imputing missing values, and aggregating player statistics to prepare data for modelling.
- Used pandas, Matplotlib, Scikit-Learn, and conducted hyperparameter optimization using GridSearchCV.