

LINGYU ZHOU

lz568@cornell.edu | 608-234-0062 | zhoulingyu.net | Ithaca, NY

EDUCATION

Cornell University - <i>Master of Professional Study in Applied Statistics, Data Science focus</i> GPA: 3.948	2023.8 - 2024.12 <i>Ithaca, NY</i>
University of Wisconsin-Madison - <i>Bachelor of Science in Mathematics, Data Science Minor</i> GPA: 3.898	2020.9 - 2023.5 <i>Madison, WI</i>

PROFESSIONAL SKILLS

Languages	Python, SQL, Stata, R, SAS 9.4 (base, certified), HTML, JavaScript, Markdown
Software & Tools	Microsoft Office, MongoDB, Apache Hive, Apache Hadoop, PowerBI, AWS, Git regex, selenium, beautifulsoup, Linux, \LaTeX , WordPress, Cytoscape, jQuery
Professional Knowledge	Regression (Linear, Multinomial, Logit, Probit, Lasso, Ridge), Transformer Perceptron, SVM, KMeans, KNN, Decision Tree, GMM, ARIMA, PCA, OLS

WORK EXPERIENCES

Cornell IRL School <i>Data Analyst</i>	2024.5 - TBD <i>Ithaca, NY</i>
--	-----------------------------------

- Reduced database size by **27%** from using transformer model BERT and json indexing along with customized query function to agglomerate similar records and remove unnecessary fields
- Leveraged Python and Stata to organize and clean panel data about labor compliance surveys across 7 years & countries with **more than 1 million rows and 20 fields**

New York Sea Grant <i>Data Analyst/Assistant</i>	2024.4 - 2024.6 <i>Ithaca, NY & Remote</i>
--	---

- Scraped data of **more than 2 GiB** from ARCGIS using GIS platform, AirTable & Python beautifulsoup packages, etc.; Resolved database mismatch issue by rewriting data structures that reduced **8% shorter time** for uploading.
- Performed **QA/QC tasks** to identify and rectify **200+ rows** erroneous flood incidents reports user uploaded; Conducted daily database maintenance and statistical analyses about flooding panel data.

Jiangsu HopeRun Software Co., Ltd. <i>Data Engineer</i>	2023.7 - 2023.8 <i>Nanjing, China</i>
---	--

- Deployed **Apache Flink local cluster** and utilized Apache Flink's APIs to process IoT batch data over 1000+, stream data simulated by TCP/IP, netcat functionalities, and MQTT environment, **57%** faster than traditional SQL.
- Conducted **market research** on SPD systems in China's smart healthcare market and contributed to the research report, identifying **over 20** items in SWOT analysis.

Shanghai SINO Environmental Technology Co., Ltd. <i>Data Analyst/Engineer</i>	2021.5 - 2022.5 <i>Shanghai, China & Remote</i>
---	--

- Supervised and processed environmental monitoring data, achieving **89%** accuracy rate and flawless operations.
- Sorted, classified and arranged processed data using Excel and Python to reduce data retrieval times by **24%**.
- Analyzed, identified and reported timely **100+** possible problems ; modeled to discern abnormal data from noises

RESEARCH EXPERIENCE

China's Live Hog Futures Market Research <i>Undergraduate Research, directed by Prof. Martin Cherkes (Princeton University)</i>	2022.7 - 2022.9
---	-----------------

- Conducted research regarding topics of securities and market analysis using statistical analysis, Python and Excel; conducted a final report presentation regarding the research topics and paper
- Refined and published an analytical **academic paper** (https://doi.org/10.1007/978-981-99-6441-3_96) on China *AEPS* (ISSN 2731-4006) and *AEMPS* (ISSN 2754-1169)

DRP Research Program <i>Undergraduate Research, directed by Jacob B Fiedler (University of Wisconsin-Madison)</i>	2022.1 - 2022.5
---	-----------------

- Researched on Computability & Kolmogorov Complexity under a PhD mentor, involving **10-hours** weekly reading
- Discussed, & presented findings weekly with tutors & peers to finally delivered a **30-minute** defense presentation