

LINGYU ZHOU

Ithaca, NY, 14850 | 608-234-0062 | lz568@cornell.edu | linkedin.com/in/lingyuzhou | zhoulingyu.net

EDUCATION

Cornell Univeristy *College of Computing & Information Science*
Master of Professional Study
Applied Statistics | Data Science Focus

Aug 2023 - Dec 2024
GPA: 3.948

Univeristy of Wisconsin-Madison *College of Letters & Science*
Bachelor of Scinece
Mathematics | Data Science Minor

Sep 2020 - May 2023
GPA: 3.898

ACADEMIC ACHIEVEMENTS

Dean's list-2021, 2022, 2023 (University of Wisconsin-Madison)
Distinctive Scholastic Achievement (University of Wisconsin-Madison)
Distinction of top 15% distinction in CEMC (12)

PUBLICATIONS

Purchasing Practices and Labor Outcomes in Global Supply Chains

Nov 2024

- Sarosh Kuruvilla, Chunyun Li, **Lingyu Zhou**
- **Under review*, British Journal of Industrial Relations

Basic Analysis of Hog Futures Market in China

Oct 2023

- Chen Xu, **Lingyu Zhou**, Anshi Lou, Manqi Qiu, Mengyu Cao
- doi.org/10.1007/978-981-99-6441-3_96, AEPS (ISSN 2731-4006), AEMPS (ISSN 2754-1169)

RESEARCH & PROJECTS

Data Analysis on East Asia Labor Compliance Rate Research Program

May 2024 - Dec 2024

Summer Research under Professor Sarosh Kuruvilla (Cornell ILR School) and Chunyun Li(LSE)

- Utilized Python and R to systematically organize and clean panel data from labor compliance surveys spanning 7 years and multiple countries, involving over **1 million** rows and 20 variables.
- Performed **multi-level regression** and **Difference-in-Differences (DiD)** analysis, to extract meaningful insights from complex panel data and implemented data visualizations including heat maps, swarm plots,
- Leveraged transformer model **BERT** to aggregate similar records and eliminated redundant fields, optimized data storage by reducing database size by **31%** with self-made data structure JSON indexing, combined with a customized query function.

China's Soybean Import Modeling and Time Series Analysis

Jan 2024 - May 2024

Professional Capstone Project under Gro Intelligence

- Conducted data mining, cleaning, & engineering to ensure data reliability of for modeling.
- Developed, and implemented advanced statistical models (MLR, AR, ARMA, ARIMAX, Random Forest) to predict China's soybean import volumes and achieve **42%** less MAPE than the simple MLR model.
- Evaluated and optimized model performance using different metrics for robust accuracy and reliability in forecasting outcomes.
- Interpreted complex data visualizations to identify trends, and correlations to provide real-world actionable insights for global agricultural markets.

China's Finance and Marketing Analysis Research

Jul 2022 - Sep 2022

Undergraduate Research, directed by Prof. Martin Cherkes (Princeton University)

- Studied under Prof. Cherkes regarding topics of securities and market analysis.
- Met for an additional **10** hours per week with team members for the research paper, including topics settlement, data mining, model selection, typesetting changes, etc.
- Held a final **30-minute** presentation on the topic and published a **20-page** refined academic paper about Live Hog Futures' Analysis in China.

DRP Research Program

Undergraduate Research, directed by Jacob B Fiedler (**University of Wisconsin-Madison**)

Jan 2022 - May 2022

- Followed a PhD mentor's research regarding topics of **Computability and Kolmogorov Complexity**.
- Conducted **10+** hours of reading per week; met with tutors and 2 other peers 1 hour a week for a semester.
- Discussed with the mentor and peers about problems that emerged and made presentations on research findings every week.
- Accomplished a **20-minute defense** presentation at the end of the semester.

TEACHING EXPERIENCES

STSCI 5740 - Data Mining and Machine Learning

Teaching Assistant, under Dr Nayel Bettache

Aug 2024 - Dec 2024

Cornell University

- Set rubrics and graded exams and **500+** students' assignments in total on Gradescope with clarified explanations.
- Answered **50+** course questions from students timely on Canvas discussion board.
- Helped set up virtual office hours, exam rooms, and classroom environments.

WORK EXPERIENCES

New York Sea Grant

Data Analyst Assistant

Apr 2024 - Jun 2024

Ithaca, NY & Remote

- 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.

Data Engineer

Jul 2023 - Aug 2023

Nanjing, China

- 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,

Data Analyst/Engineer

Jul 2020 - Jun 2021

Shanghai, China & Remote

- 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

PROFESSIONAL SKILLS

Languages

Software & Tools

Professional Knowledge

Python, SQL, Stata, R, Base SAS 9.4 (certified), HTML, JavaScript, Matlab
Microsoft Office, MongoDB, Apache Hive, Apache Hadoop, jQuery
L^AT_EX, PowerBI, AWS, Git regex, selenium, beautifulsoup, Linux, Wordpress
Regression (Linear, Multinomial, Logit, Probit, Lasso, Ridge), GMM, ARMA
Transformer Perceptron, SVM, KMeans, KNN, Decision Tree, PCA, OLS

RELEVANT COURSES

Core Courses

Multivariable Calculus
Advanced Linear Algebra & Differential Equations
Nonlinear Dynamics & Chaos
Mathematical Analysis
Mathematical Methods in Data Science
Probability & Statistics Theory
Database management and Big Data Analysis

Other Courses

Data Science Software Engineering
Machine Learning & Reinforcement Learning
Programming: Java, Python
Advanced Econometrics
Micro/Macro economy
Portfolio Management and Operational Research
Computational Biology/Bio-statistics