Yongchan Lee

+1 330-347-4890 • yongchanleee@gmail.com • linkedin.com/in/yongchanleee • ianyongchanlee.github.io

TECHNICAL SKILLS

Programming & Query Languages: Python (Pandas, Scikit-learn), SQL-MySQL, R
Data Analysis & Statistical Methods: Exploratory Data Analysis (EDA), Regression Analysis, Time-Series Analysis, Data Validation
Data Visualization & Reporting: Tableau, Excel
Data Collection & Preparation: API-Based Data Ingestion, Data Cleaning, Data Wrangling, Feature Engineering, Automated Data Pipelines

EXPERIENCE

Sherwin-Williams Wooster, OH
Intern, Data Analyst May 2025 - Dec 2025

- Translated an ambiguous risk forecasting request into a clearly defined KPI by formalizing risk as expected financial loss, integrating machinery performance and financial datasets.
- Resolved data granularity constraints through data mapping, recovering 30% of unusable data to enable work-center-level analysis.
- Built and evaluated a Python- and SQL-based risk analysis model, achieving 90% confidence (±5%) in risk tier prediction.
- Supported operational decision-making by classifying work centers into high, medium, and low risk tiers and presenting comparable analytical scenarios to stakeholders.

Ulsan National Institute of Science and Technology (UNIST) Ulsan, Korea
Intern, Research Analyst Jun 2024 - Aug 2024

- Built an automated Python pipeline using NASA APIs to ingest, clean, and integrate large-scale environmental datasets, enabling scalable multi-dimensional analysis.
- Designed an analysis-ready dataset by engineering environmental and vegetation features aligned with crop growth stages.
- Validated the analytical value of the engineered features through regression modeling, achieving a 0.15 R^2 improvement over published benchmarks

The College of Wooster Wooster, OH
Teaching Assistant Aug 2025 - Dec 2025

- Supported instruction in regression, classification, and statistical inference, emphasizing assumptions, interpretation, and data integrity.
- Guided students through end-to-end data analysis workflows, including data cleaning, exploratory analysis, visualization, and result validation.

PROJECT

Climate-Driven Crop Yield Analysis| *Honor Awarded* Jan 2025 - Dec 2025
Senior Thesis

- Reduced data preparation time by 50% through an automated pipeline that ingested and integrated 5 heterogeneous sources, standardizing disparate temporal and regional resolutions into unified county-level analytical tables.
- Built 0.75-0.8 R^2 regression-based yield models for three major crops to analyze relationships between climate variables and crop productivity.
- Pinpointed crop-specific drivers by identifying the exact monthly climate variables that most significantly impacted yield at each growth stage.
- Investigated underperforming wheat model through anomaly analysis and supplemental research, developing testable hypotheses to explain weak correlations and inform future analytical refinement.

Life Expectancy Analysis Jan 2025 - May 2025

- Conducted comparative analysis using FAO and World Bank datasets to evaluate relationships between dietary patterns and life expectancy.
- Developed an interactive Tableau dashboard using multiple visualization types to communicate key dietary and health insights across countries.
- Tested and ruled out geographic bias by controlling for country location, confirming that observed differences in life expectancy were driven by dietary patterns rather than regional effects.
- Validated analytical findings through literature review, confirming that higher intake of specific nutrients identified in the analysis is scientifically associated with increased life expectancy.

EDUCATION

The College of Wooster Wooster, OH
B.A. Major: Statistical & Data science (Department Honor) | Minor: Environmental Studies *Aug 2020 - Dec 2025*
Cumulative GPA: 3.913 / 4.0