

# Yongchan Lee

+1 330-347-4890 • [yongchanleee@gmail.com](mailto:yongchanleee@gmail.com) • [linkedin.com/in/yongchanleee](https://www.linkedin.com/in/yongchanleee) • [ianyongchanlee.github.io](https://github.com/ianyongchanlee)

## SUMMARY

Experience turning messy, multi-source data into validated, analysis-ready datasets and KPI reporting layers. Skilled in Python, SQL, and Tableau, with strengths in data quality checks, metric definition, and stakeholder-ready insights.

## TECHNICAL SKILLS

**Programming & Query:** Python (Pandas, Scikit-learn), SQL (MySQL, Postgres), R-studio

**Data Analytics:** 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 Pipelines

## EXPERIENCE

### Sherwin-Williams

Wooster, OH

*Intern, Data Analyst*

May 2025 - Dec 2025

- Reframed an ambiguous risk request as a measurable KPI by formalizing risk as expected financial loss.
- Resolved data granularity constraints with Excel through data mapping, recovering 30% of unusable data to enable work-center-level analysis.
- Built and validated a Python risk-tier model from SQL-extracted operational features, achieving 90% classification accuracy.
- Informed ERM's mitigation planning by ranking sites/work centers into High/Medium/Low risk tiers and providing comparable scenarios, guiding where to focus controls to prevent future loss events.

### 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.
- Designed an analysis-ready dataset by engineering environmental and vegetation features aligned with crop growth stages.
- Validated analytical value of engineered features through regression analysis, achieving 0.15 improvement in  $R^2$  over published benchmarks.

## PROJECT

### Manufacturing Analytics Warehouse & KPI Mart (PostgreSQL)

- Built a PostgreSQL star-schema warehouse integrating IoT machine signals and ERP outputs into a consistent Line  $\times$  Day reporting layer.
- Converted RUN/STOP state logs into daily runtime/downtime by reconstructing operating sessions and handling incomplete or noisy timestamps, then materialized a KPI mart (availability, yield, throughput).
- Improved query speed and decision-ready reporting with indexing and an analysis/QA layer (downtime Pareto, yield/throughput trends)

### Climate-Driven Crop Yield Analysis

- Automated data ingestion from 5 heterogeneous sources, reducing manual preparation time by  $\sim 50\%$ .
- Standardized disparate temporal and regional resolutions into unified county-level analytical tables.
- Built regression-based yield models ( $0.75\text{--}0.8 R^2$ ) to analyze relationships between climate variables and crop productivity.
- Identified crop-specific monthly climate drivers and investigated underperforming segments through anomaly analysis.

### Life Expectancy Analysis

- Integrated FAO food availability (food groups + nutrient indicators) with World Bank health metrics using country codes/years.
- Developed a Tableau dashboard to explore food groups vs nutrients, and tested insights via top-5 vs bottom-5 life expectancy comparisons.
- Reported actionable insights: moderate correlation for energy/protein and a pronounced fish consumption gap in high-life-expectancy cohorts.

## 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