



Jinwook JUNG's Portfolio

2026 summer Internship application

Profile & Positioning

Jinwook (Alex) Jung

Data Scientist

Purdue University

B.S. Data Science & Applied Statistics | Minor in Economics

Expected Graduation: **Dec 2026**

Positioning Statement

As a data scientist, I prioritize analysis driven by accuracy, reproducibility, and actionable decision-making to support reliable business outcomes.

Key Domains

- Medical & Healthcare Analytics
- Scientific Data Engineering
- Public & Global Data Analysis



Project 1. AI-Omics Lab (Purdue University)

Scientific Data Engineering & Backend Validation

Context

- Member of the AI-Omics Research team under Dr. W. Andy Tao's Lab
- Disease biomarker ML research using **high-resolution mass spectrometry (Mass Spectrometry)** data

Problem

- Large-scale scientific datasets with **inconsistent structures across experiments**
- Unvalidated input data prevents reliable ML model training
- **Reproducibility** is a critical constraint in biomedical research

My Contribution (Core Work)

- Backend data pipeline (Python): **File I/O, modular code design, exception handling**
- Developed **rule-based validation logic** to automatically verify data integrity before downstream analysis
- Performed **data preprocessing and standardization** to ensure consistent data structures prior to ML training
- Collaborated in **backend development**, gaining exposure to production-style backend architecture
- Used **Git/GitHub** for collaborative development, including feature branching and rebasing workflows

Output

- Designed and implemented **backend validation logic** to ensure the integrity of mass spectrometry data, enabling a stable preprocessing pipeline and improving the **reproducibility of AI-driven omics research**.

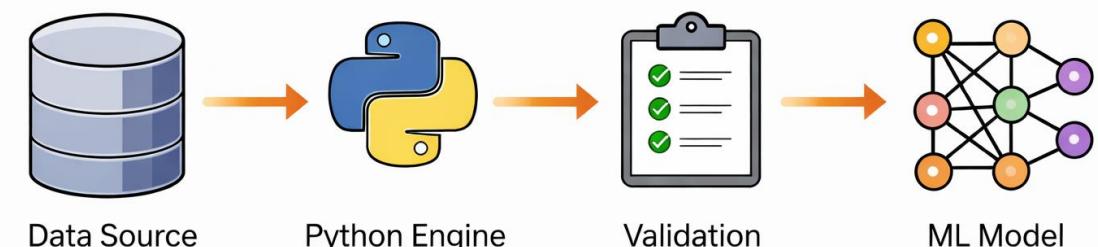
Skills Used

- Python (Core, Backend Logic)
- Django, Data Validation
- Git/GitHub, Collaborative Development

Why It Matters

“Gained a practical understanding that **ensuring data reliability before model** development is one of the most critical data science skills in real-world research environments”

Backend Architecture Diagram



Project 2 University Hospital Limerick (Ireland)

Healthcare & Epidemiological Data Analytics

Context

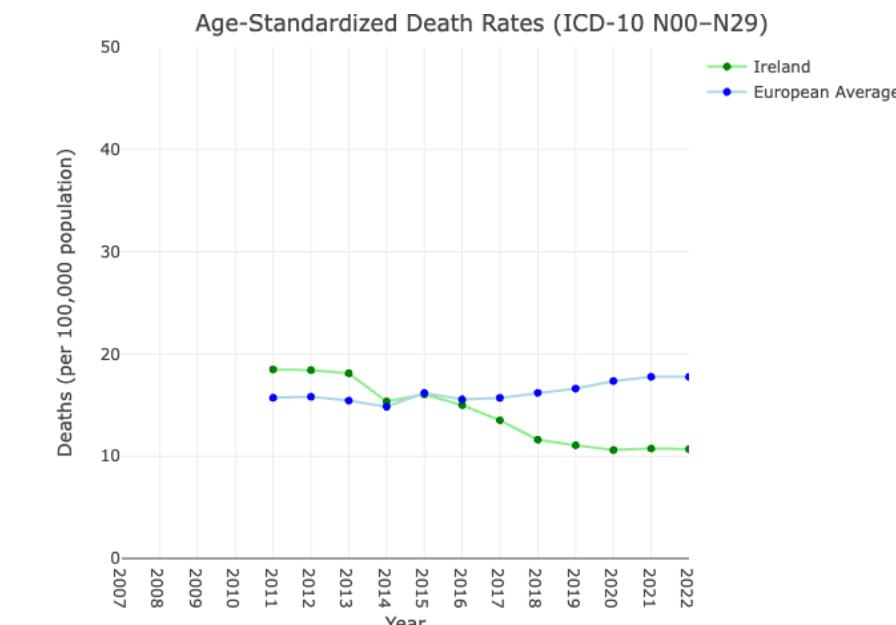
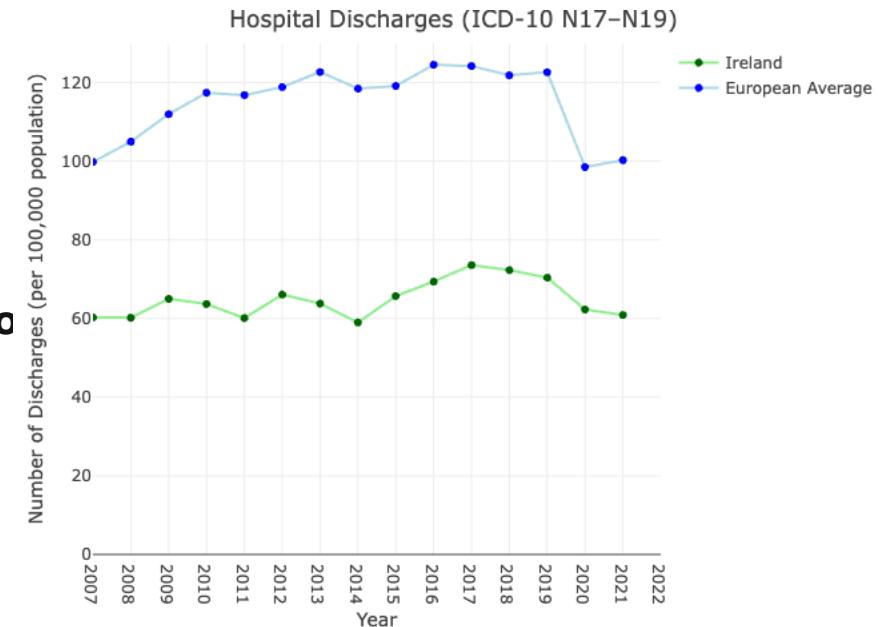
- Analysis of **15+ years of healthcare data** with **cross-country comparisons across 33 EU countries**

Problem

- **Heterogeneous healthcare data structures** across countries and regions
- Required **public-health-oriented interpretation**, beyond simple descriptive statistics

My Contribution

- **Data Wrangling & Integration**
 - ✓ Utilized tidyverse packages (dplyr, tidyr,forcats)
- **Epidemiological Analysis**
 - ✓ Applied Age Standardization and Age-Standardized Death Rates (ASDR)
 - ✓ Classified disease outcomes based on **ICD-10** coding standards
- **Geospatial & Interactive Analytics**
 - ✓ sf, leaflet, giscoR
 - ✓ Built **interactive dashboards** using **Plotly** and **Crosstalk**



- **Reproducible Research**
 - ✓ Quarto based analytical reports

Output

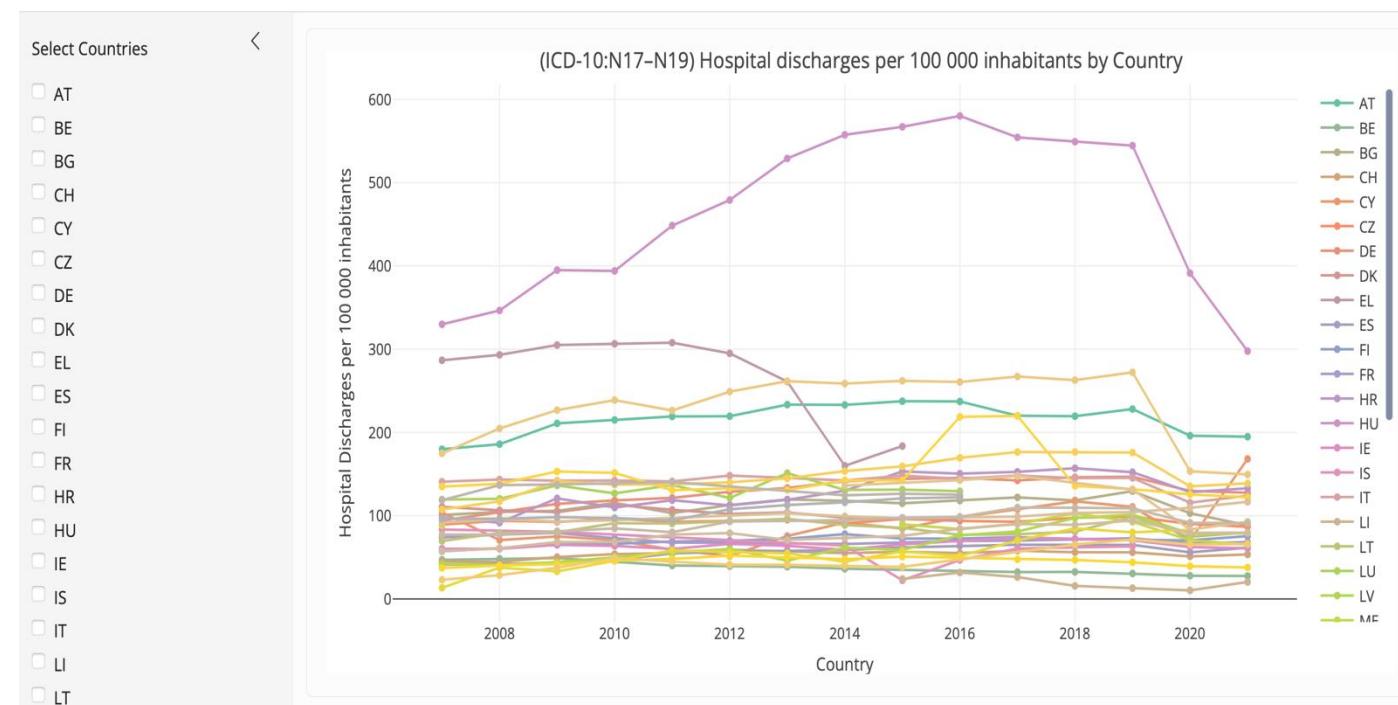
- Time-series and map visualizations of **hospitalization and mortality rates** based on selected EU countries
- Supported **data accessibility** and **decision-making** of the research

Skills Used

- R, Tidyverse, Epidemiology, GIS, Dashboarding

Why It Matters

- Converted **complex healthcare data into accessible formats**
- Enabled **cross-country comparisons** of healthcare indicators using a consistent standard
- Provided dashboards that **directly support research and decision-making**



3. Other Data Operations Experience

Applied Data & Operational Experience

ACT Foundation (Seoul)

- Managed an asset database with **1,000+ records**
- Improved **data accessibility** and **operational efficiency**
- Gained hands-on experience in **data operations within a nonprofit organization**

Key Takeaways

“Developed an understanding of practical, real-world data management and operations”



FINANCIAL INNOVATION



Data Science Fit Why My Background Fits Data-Driven Roles

1. 데이터 신뢰성 우선순위

정확성과 검증을 중시하는 의료 데이터 경험.

2. 의사결정 중심 분석

단순 모델링보다는 해석적 분석에 중점을 둔 연구 및 공공 데이터 작업 경험.

3. 글로벌 및 분야 간 경험

한국, 미국, 유럽의 데이터에 대한 노출 경험.

의료 데이터에서 보험, 리스크 관리, 금융 데이터로 전문성 확장 가능성.

“저는 데이터를 통해 신뢰할 수 있는 의사결정 프로세스를 설계하는 Data Scientist가 되고자 합니다.”