# Yejin Hwang

+1 361-800-2613 | yejincc99@gmail.com | linkedin.com/in/yejin-data | github.com/Yejin-Hwang

## Education

**Texas A&M University–Corpus Christi** — Corpus Christi, TX
*Master of Data Science; GPA: 4.0* — *Expected Dec 2025*

- Foundation Models (Time-Series / Language), Time-Series Forecasting

**Sungkyunkwan University (SKKU)** — Seoul, South Korea
*B.A. in Culture & Technology (Data Science); GPA: 3.72* — *Aug 2023*

## Technical Skills

**Languages**: Python, R, SQL
**Frameworks**: Scikit-learn, TensorFlow, PyTorch, Streamlit
**Tools**: Git, AWS, MySQL, Tableau, Power BI, Excel, VS Code, Jupyter Notebook, LaTeX
**Libraries**: Pandas, NumPy, Scipy, Yfinance, Matplotlib, Seaborn, Transformers, Statsmodels

## Professional Experience

**Research Assistant** — Sep 2024 – Present
*Texas A&M University-Corpus Christi* — *Corpus Christi, TX*

- Led research on time-series forecasting of stock data (TSLA, AAPL, NVDA) using ARIMA, TimesFM, and Chronos-T5 models, targeting short-term (1–7 days) forecasts.
- Engineered key features including lag variables, rolling statistics, and volatility measures to enhance forecast accuracy.
- Improved model performance by 30% MAE reduction, potentially achieving a $2,000 gain from a hypothetical $100,000 portfolio investment.
- Built reproducible Python pipelines integrating yfinance API and PyTorch for data ingestion, model training, and performance reporting to support financial decision-making.
- Presented model comparison and forecasting results at Joint NMSU/UTEP Conference to over 30 attendees, highlighting real-world financial implications.

**Data Research Associate** — Aug 2022 – Mar 2024
*Samsung Medical Center* — *Seoul, South Korea*

- Collected and analyzed patient hair data from breast cancer patients using Folliscope 5.0, evaluating chemotherapy-induced alopecia severity with high precision.
- Collected, cleaned, and validated CRF data for esophageal, breast, and lung cancer patients, improving data quality and analysis readiness by 30%.
- Developed comprehensive codebooks and data dictionaries for standardized data documentation, enhancing reproducibility and collaboration across research teams.
- Collaborated with clinicians to deliver reliable datasets, directly supporting ongoing cancer research and informing clinical decision-making.
- Improved data integrity, contributing to enhanced patient care strategies and more efficient treatment planning at Samsung Medical Center.

## Data Science Projects

**Bayesian Healthcare Risk Modeling (Finance-Focused Capstone)** | *R, MCMC(HMC), Regression* — May 2025

- Developed probabilistic regression model for forecasting insurance charges, supporting actuarial cost modeling and financial risk assessment.
- Applied Bayesian hierarchical regression with nonlinear effects (age$^2$) and BMI×smoker interaction; validated via HMC and WAIC.
- Outperformed frequentist models (RMSE ↓2.7%, MAE ↓2.1%); enabled better uncertainty quantification for data-driven financial policy design.

## Presentations & Research

**Joint NMSU/UTEP Conference on Math, Computer Science & Computational Sciences** Apr 2025
*Presented forecasting model comparison using Chronos-T5 and TimesFM on TSLA data* *Las Cruces, NM*

**Coastal Bend Mathematics and Statistics Conference** Apr 2025
*Presented research on transformer vs. traditional models for stock price prediction* *Corpus Christi, TX*

## Awards

**SKKU Dean's Award for Excellence in Global Engagement** Mar 2023
*Recognized for top-level academic achievement and cross-cultural research initiatives* *Seoul, South Korea*