

FRANCIS PAUL C. FLORES

Data Scientist | Healthcare & Market Research Analytics

francispaulflores6@gmail.com | kaggle.com/grosvenpaul/code (Peak Rank: #158 globally) | www.francispaulflores.com

PROFESSIONAL SUMMARY

Data Scientist with 9+ years of experience spanning academic healthcare research and market analytics, specializing in statistical modeling, ML pipeline development, and analytics automation. Coauthor on peer-reviewed publications in The Lancet, Nature Medicine, and JAMA Network Open; recipient of the HKU Knowledge Exchange Excellence Award (2024) for translating complex population health analyses into policy-relevant insights. Proficient in Python and R across the full data lifecycle, from feature engineering and supervised/unsupervised ML to time series forecasting and automated reporting.

CORE COMPETENCIES

Programming & Tools: Python, R, SQL, SAS, STATA, SPSS

Machine Learning & AI: Supervised/Unsupervised Learning, LLM Fine-Tuning, Random Forests, Gradient Boosting, NLP Pipelines

Time Series & Forecasting: ARIMA, VAR, GARCH, Prophet, LSTM, Interrupted Time Series (ITS)

Statistical and Longitudinal Modeling: GEE, Mixed-Effects Models, Latent Class Mixed Models, Survival Analysis

Market Research Methods: MaxDiff Scaling, Shapley Value Regression, Consumer Preference Modelling, Segmentation

Data Engineering & Automation: R/Python Pipeline Development, RMarkdown, Automated Reporting, Reproducible Workflows

Visualization & Reporting: ggplot2, Shiny, Plotly, Seaborn

Libraries: tidyverse/dplyr, data.table, pandas, NumPy

EXPERIENCE

Data Scientist

December 2025 – Present

Human8 | Philippines

- Engineered an automated Python and Plotly dashboard suite for key tracker studies, reducing client reporting turnaround by over 95% and saving analysts 20 hours weekly in manual reporting bottlenecks.
- Fine-tuned an XLM-RoBERTa-Large AI model to categorize hundreds of thousands of open-ended survey responses against established tracker codeframes; achieved an ~80% agreement rate with human coders while compressing turnaround time from weeks to hours and substantially reducing operational costs across qualitative data processing workflows.

Senior Technical Officer

November 2021 – December 2025

Technical Officer

November 2019 – November 2021

Research Assistant II

March 2018 – November 2019

School of Public Health, The University of Hong Kong | Pok Fu Lam, Hong Kong SAR

- Architected reproducible R/Python analytical pipelines processing 5M+ rows of Hong Kong government data, automating EDA, feature engineering, model training, and reporting — standardizing the team's end-to-end research workflow and eliminating manual, one-off scripts.
- Quantified population-level health risk factors across Hong Kong's 18 districts by applying supervised ML (GLM, Random Forests, SVM, Naive Bayes) and unsupervised segmentation (K-Means, Latent Class Analysis), directly informing resource allocation and intervention planning for the Food and Health Bureau.
- Modeled public health trends, intervention effects, and longitudinal risk trajectories using time series methods (ARIMA, VAR, GARCH, Prophet, LSTM) and panel models (mixed-effects, GEE), underpinning multiple publications across leading journals reaching millions of policymakers and clinicians worldwide.
- Authored statistical methods and results sections for manuscripts in The Lancet, The Lancet Public Health, Nature Medicine, JAMA Network Open, and The Lancet Regional Health, translating model outputs into clear, evidence-based narratives for global health and policy audiences.
- Recognized with the HKU Knowledge Exchange Excellence Award (2024) for bridging academic research and real-world public health policy impact.
- Collaborated on study design, power calculations, and analysis plans; mentored junior researchers and students in statistical modeling and reproducible coding best practices.

Assistant Statistician

May 2016 – December 2017

Goodthinking Research Inc. | Pasig City, Metro Manila, Philippines

- Championed the adoption of R across the analytics team (migrating from XLStat), establishing standardized RMarkdown workflows that boosted reporting reproducibility and reduced manual effort across the team.
- Implemented MaxDiff Scaling and Shapley Value Regression to quantify consumer preference drivers, delivering statistically rigorous, client-ready market research insights for product and brand strategy decisions.

- Built predictive models, demand forecasts, and latent consumer segmentations to identify market demand patterns, directly informing clients' go-to-market and product positioning strategies.

SELECTED PUBLICATIONS

- Ni, M. Y., et al., & Flores, F. P., et al. (2020). Depression and post-traumatic stress during major social unrest in Hong Kong: a 10-year prospective cohort study. *The Lancet*, 395(10220), 273–284. [https://doi.org/10.1016/S0140-6736\(19\)33160-5](https://doi.org/10.1016/S0140-6736(19)33160-5)
- Ni, M. Y., Canudas-Romo, V., Shi, J., Flores, F. P., et al. (2021). Understanding longevity in Hong Kong: a comparative study with long-living, high-income countries. *The Lancet Public Health*, 6(12), e919–e931. [https://doi.org/10.1016/s2468-2667\(21\)00208-5](https://doi.org/10.1016/s2468-2667(21)00208-5)
- Shi, J., Leung, C. M. C., Chen, R., Xiao, X., Flores, F. P., et al. (2026). Interpersonal conflicts, social media use and depression associated with protests. *Nature Medicine*, 32(1), 224–230. <https://doi.org/10.1038/s41591-025-04145-0>
- Lun, P., Ning, K., Wang, Y., Flores, F. P., et al. (2023). COVID-19 Vaccination Willingness and Reasons for Vaccine Refusal. *JAMA Network Open*, 6(10), e2337909. <https://doi.org/10.1001/jamanetworkopen.2023.37909>
- Ni, Y., Flores, F. P., Lun, P., Ning, K., et al. (2024). Understanding paediatric COVID-19 vaccination during the pandemic: a prospective cohort and population-based registry study. *The Lancet Regional Health – Western Pacific*, 43, 100976. <https://doi.org/10.1016/j.lanwpc.2023.100976>

EDUCATION

Bachelor of Science in Statistics

Mandaluyong City, Metro Manila, Philippines

Rizal Technological University

- President, Organization of Statistics Students | Leadership Excellence Award

RECOGNITION & AWARDS

- ▶ HKU Knowledge Exchange Excellence Award (2024) hku.hk/award/ceremony/award-recipients-2024/953
- ▶ Kaggle Notebooks Expert — Peak Rank: #158 globally kaggle.com/grosvenpaul/code
- ▶ Kaggle Weekly Kernel Award — “EDA and Time Series Modeling” (2017) twitter.com/kaggle/status/923980305291251714
- ▶ Invited Resource Speaker: “Predictive Analytics, Forecasting the Future” Rizal Technological University, Mandaluyong City

CONTINUOUS LEARNING

Completed MOOCs: Introduction to R for Data Science (Microsoft) • The Analytics Edge (MITx) • Machine Learning (Stanford / Andrew Ng) • MIT's Open Learning Data Science Curriculum

Actively applies emerging methodologies in Deep Learning and NLP, as demonstrated by the deployment of XLM-RoBERTa workflows and a public portfolio of advanced Kaggle notebooks.