# Yama (Ya-Wen) Chang

Tel: 646-927-9876 | Email: yamachang.ds@gmail.com | LinkedIn | Portfolio | New York, NY

## EXPERIENCE SUMMARY

Data Scientist with over 5 years of experience focusing on survey-based research, analytics, and data-driven insights. Specializes in experimental design, psychometrics, and modeling with a strong understanding of human behavior.

## PROFESSIONAL EXPERIENCE

**Data Scientist,** ***Lab for Scalable Mental Health*** | *Stony Brook, NY* ***August 2022 – Present***

- Enhanced solution efficiency by 90% through the implementation of data-driven analytical tools and automation.
- Collaborated and communicated with cross-functional partners to enhance decision-making using data analytics, survey research, experimentation, and modeling.
- Identified key metrics and performed rigorous exploratory and statistical analysis to understand demographic, behavioral, and attitudinal trends, leading to published papers in high-ranking journals.
- Led and promoted best practices for data management, including data quality checks, data cleaning, data dictionary creation, and statistical analysis.

**Senior Data Analyst,** ***University of Pittsburgh Medical Center*** | *Pittsburgh, PA* ***June 2020 – July 2022***

- Designed and implemented research studies of survey and behavioral data, yielding critical insights into target demographics and leading to published papers in high-ranking journals.
- Automated data aggregation from multiple sources, leading to a 20-hour monthly reduction in feature engineering and transformation processes.
- Created a markdown-based reporting system for visualizing recruitment indices (e.g., Year-over-Year Growth), saving an additional 20 hours per month.

## TECHNICAL SKILLS

- **Programming:** R (tidyverse, ggplot2, tidymodels, markdown), Python (numpy, pandas, scikit-learn, matplotlib)
- **Tools:** SQL, Qualtrics, Git, GitHub, Command Line
- **Stats & Machine Learning:** Covariance and correlation modeling, Hypothesis Testing, A/B Testing, Multilevel modeling, Factor Analysis, Principal Component Analysis, Cluster Analysis/Latent Profile Analysis, Classification (KNN, Random Forest, XGBoost), Regression (Lasso, Ridge, Elastic Net), Clustering (K-means, GMM) Modeling

## SELECTED PROJECTS

***Real Estate Prices Prediction*** ***2023***

*Machine Learning Regression | Python*

- Developed a predictive model using machine learning algorithms, including Regression Models, Random Forest, and Gradient Boosting Decision Trees, to estimate real estate prices.
- Optimized model performance through cross-validation and hyperparameter tuning, achieving a Mean Absolute Error under the pre-set benchmark ($70k), reflecting superior prediction accuracy.

***Geospatial Analysis of Structural Stigma towards LGBTQ+ youth in the U.S.*** ***2023***

*Factor Analysis & Regression Analysis | R*

- Extracted, cleaned, and analyzed large open datasets (600000+ raw data) to conduct a comprehensive analysis of attitudes towards LGBTQ+.
- Created and visualized a county-level discrimination index using principal component analysis.

***Identify Long-Term Suicidal Ideation Profiles in Late-Life Depression*** ***2022***

*Cluster Analysis | R*

- Utilized cluster analysis to identify distinct profiles over time in elderly individuals with major depressive disorder.
- Handled a large dataset of 468 participants across 16 years, showcasing proficiency in managing complex data.
- Visualized distinct profiles using radar plots, enhancing the understanding of diverse patterns of suicidal thoughts.

## EDUCATION

**Columbia University** | *Master of Art* ***2018 - 2020***

Clinical Psychology with a concentration in Research/Quantitative Methods (GPA: 4.0/4.0)

Coursework: Data Science; Machine Learning; Probability and Inferential Statistics; Applied Regression Analysis

**Stanford University** | *Exchange Program* ***2010***

**National Taiwan University** | *B.A. in Economics* ***2008 - 2012***