Justin Xu

New York, NY • justinxiaonixu@gmail.com • (773) 717-4113 • https://www.linkedin.com/in/justinxxu/

Healthcare Data Scientist (MS Biostatistics) with experience in predictive modeling, clinical data analysis, and analytics product development. Strong foundation in statistical methods including causal inference and survival analysis, combined with Python/ R programming skills for building ML pipelines and data-driven solutions. Proven ability to translate complex healthcare data into actionable insights through dashboards, statistical reports, and automated workflows. Experienced in clinical trial analytics , patient risk stratification, and real-world evidence generation.

SKILLS

Clinical & Statistical Tools: Python (pandas, numpy, scikit-learn, matplotlib, seaborn, Jupyter), R (tidyverse, ggplot2, RShiny , R Markdown, MatchIt, survival), SQL, Git/GitHub, Data Pipelines , Statistical Analysis Plan (SAP), SAS.

Machine Learning & Analytics: Predictive Modeling (Classification, Regression, Model Evaluation), Statistical Methods (Causal Inference, Survival Analysis, GLM, ANCOVA, PCA), Healthcare Analytics (EHR/Claims Data, Patient Risk Stratification), A/B Testing, Statistical Inference, Deep learning, supervised/unsupervised learning

Data Engineering & Visualization: Data Wrangling (Cleaning, Transformation, ETL), Interactive Dashboards (RShiny), Publication-Quality Graphics, High-Dimensional Data Processing

PROFESSIONAL EXPERIENCE

Columbia University Valeri Lab, Research Assistant, New York, NY — Apr. 2025 – Present

- Engineered automated causal mediation analysis pipeline in R (CMAverse package) processing 5,000+ longitudinal patient observations, enabling 200+ healthcare researchers to analyze time-varying mediators reduce analysis time from weeks to days
- Led development of novel multi-timepoint mediator function for CMAverse R package (345 citations, 200+ active users), delivering comprehensive testing suite and published implementation guide that accelerated healthcare research workflows

Columbia University Ogden Lab, Research Assistant, New York, NY — Oct. 2024 – Present

- Automated biomedical image processing pipeline analyzing 650+ biological objects (mitochondrial/egg morphology), replacing manual labeling workflows and enabling high-throughput quantitative analysis for disease prediction research
- Validated and transformed high-dimensional functional data (2D/3D cellular morphology) using R (fdasrvf, MHD packages), computing complex distance matrices that quantified morphological variations relevant to medical diagnostic applications , also eliminated the image artifact by 40%
- Generated publication-quality visualizations and statistical reports communicating morphological analysis findings to biomedical research teams, supporting evidence-based decision making for diagnostic biomarker development

Bristol Myers Squibb, Part-time Analyst (Adjunctive Pimavanserin), Chicago, IL — Aug. 2023 – Oct. 2023

- Built R Shiny clinical trial analytics dashboard for Phase 3 Major Depressive Disorder study (280 patients), creating interactive prototype adopted by BMS clinical operations team as validation framework for future efficacy analyses
- Analyzed pimavanserin adjunctive treatment efficacy using R statistical methods, generating publication-quality visualizations (longitudinal plots, bar charts) that communicated treatment response patterns to clinical stakeholders

IQVIA, Part-time Analyst (Efficacy and Safety on Sarecycline), Chicago, IL — Aug. 2023 – Oct. 2023

- Developed automated data cleaning and statistical analysis pipeline in R for Phase 3 acne vulgaris trial (1,000 subjects), implementing SAP-compliant ANCOVA and Cochran-Mantel-Haenszel tests that reduced FDA documentation preparation time by 50%
- Executed inferential statistical analyses (ANCOVA, CMH tests) on sarecycline efficacy data, generating evidence documentation that supported regulatory submission process and accelerated drug approval pathway

DATA SCIENCE PROJECTS

Healthcare Readmission Prediction Model [In Progress], Python — Jan. 2025 – Present

- Building end-to-end ML pipeline to predict 30-day hospital readmissions using patient demographics, comorbidities (CHF, diabetes), and prior utilization patterns on 50K synthetic patient dataset with clinically grounded feature design
- Applying Logistic Regression and Random Forest classifiers with clinical feature engineering (LACE score, Charlson Comorbidity Index); evaluating model performance via ROC-AUC and precision-recall analysis | Python, pandas, scikit-learn, matplotlib

PUBLICATION & CONFERENCE PRESENTATION

Xu, X., Natale, R. (2024). Correlated evolution of beak and braincase morphology is present only in select bird clades. Journal of Morphology, 285, e21703. https://doi.org/10.1002/jmor.21703

Xu, X. "3D morphometrics of shorebird skulls show a strong signal of semi-independent evolution of beak and braincase." Annual meeting of the American Ornithological Society. August 2021.

EDUCATION

Columbia University in the City of New York, New York, NY — Sept. 2024 – Present

Master of Science in Biostatistics

Relevant Course Topics: Statistical Inference, Statistical Learning, MCMC, GLM, Regression, Modern Analysis, Probability

University of Chicago, Chicago, IL — Sept. 2019 – Mar. 2023

Bachelor of Arts in Biology

Awards: Jeff Metcalf Fellowship Grant ($5000), Ecology and Evolution Undergraduate Research Fellowship (2x)